A High Throughput Computational Analysis of MicroRNA Signatures in Human Ovarian Cancer

By

Sheema Sameen

29-FBAS/MSBI/F09

Department of Environmental Sciences
Faculty of Basic and Applied Sciences
International Islamic University
Islamabad
2012

In the name of ALLAH, the Most Beneficent, the Most Merciful

Department of Environmental Sciences
International Islamic University Islamabad

Dated: 2-2-2012

FINAL APPROVAL

It is certificate that we have read the thesis submitted by Ms. Sheema Sameen and it is our judgment that this project is of sufficient standard to warrant its acceptance by the International Islamic University, Islamabad for the M.S Degree in Bioinformatics

COMMITTEE

External Examiner

Internal Examiner
Dr Sobia Tabassum
Assistant Professor
International Islamic University, Islamabad

Supervisor
Dr Shaheen Shahzad
Assistant Professor/Chairperson
International Islamic University, Islamabad

Supervisor
Shaukat Iqbal Malik
Associate Professor
Department of Bioinformatics
Muhammad Ali Jinnah University, Islamabad

Dean, FBAS
Dr Muhammad Irfan Khan
International Islamic University, Islamabad

A thesis submitted to Department of Environmental Sciences,
International Islamic University, Islamabad as a partial
fulfillment of requirement for the award of the
degree of MS Bioinformatics

Dedication

This thesis is dedicated to my Father Dr. Muhammad Imtiaz Zafar and to my mother Afiefa Shafqat

For their prayers, love, endless support and encouragement.

DECLARATION

This thesis is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to specify this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions. No part of the thesis has been previously presented for any other degree.

Date 2-2-2012

Sheema

Sheema Sameen

CONTENTS

ACKNOWLEDGEMENT

My special thanks to Almighty Allah for His blessings, guidance, and knowledge that He bestowed. I am very grateful to Him from the deepest core of my heart as a guider who guided me well and enables me to face the challenges during this study. All glory and honor to Him. I express thanks to Holy Prophet P.B.U.H, who is forever a model of guidance and knowledge for humanity.

It is difficult to overstate my gratitude to my supervisors, Dr. Shaukat Iqbal Malik Associate Professor, Mohammad Ali Jinnah University and Dr Shaheen Shehzad, Assistant professor /Chairperson whose enthusiastic supervision, support, stimulating suggestions and encouragement truly helped the progression and smoothness of thesis.

I am deeply indebted for the support of my companion in this project, Zoya Khalid for the help and supervision that supervisors don't have time for. With her, research is always a fun. Thank you, Zoya for always being so kind to me.

My special thanks go to my dear friend Sidrah Mahmood for her so much support and prayers.

I cannot forget to express appreciation to all of the scholars of lab Sumra Wajid Abbasi, Asma Abro, Mehrin Gul, Ambreen Qadir , Javeria Ashraf, Uzma Malik,Saba munawar, Zurah Abbasi,Fatima Subhani, Madiha Jehan Azhar , Arshia Irum, Fatima Subhani and Yusra Sajid for sharing their ideas and interests with me during thesis.

I pay my special thanks to my parents for their love and affection, they always act as a moral support for me, without their interest and cooperation, I would not have ever been able to carry out this task.

My special thanks and prayers for my younger sisters Nahdia and Arshiya who really helped me a lot by sharing my burden at home during my hard times when it was really needed.

Before I finish, I would like to thank all the respondents who shared their precious ideas and answered my questionnaires.

Sheema Sameen

LIST OF ABBREVIATIONS

MiRNA	MicroRNA
PDK	Pyruvate dehydrogenase kinase
PDC	Pyruvate dehydrogenase complex
CLDN	Claudin
TJ	Tight junctions
GEO	Gene Expression Omnibus
Mev	Multi Experiment Viewer
SAM	Significance Analysis of Microarrays
FDR	False Discovery Rate
SOTA	Self Organizing Algorithm
SAGE	Serial Analysis of Microarrays
TPM	Tags per million
GENT	Gene Expression Database of Normal and Tumor tissues

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Cancer is one of the most serious diseases worldwide. Scientists have been trying to elucidate the molecular mechanisms that cause cancer development and cancer prevention. A recently identified class of non-protein-coding small RNAs, microRNAs (miRNAs), provided new insight in cancer research. They have strong impact on the expression of protein-coding genes. They are deemed to play a crucial role in the initiation and progression of human cancer, and those with a role in cancer are designated as oncogenic miRNAs (oncomiRs). Ovarian cancer is the sixth most common cause of cancer death among woman worldwide. The aim of this study was to reveal the regulated gene associated microRNA's in human ovarian cancer. Gene expression microarrays hold tremendous amount of biologically significant data. In order to systematically investigate the associated genes from this bulk of expression data various statistical and clustering techniques were explored. Statistical techniques include t-test, significance analysis of microarrays and self organizing tree algorithm (SOTA), while among clustering techniques hierarchical and K-means clustering was done. Then the most susceptible gene families were further analyzed for their potential microRNA targets by using various prediction tools including miRGen, pictar, miRanda etc.

This transcriptome analysis revealed the major representation of two gene families in ovarian cancer the pyruvate dehdrogenase kinase and claudin gene family. All of the members of these families except cldn5 were found upregulated in ovarian cancer and hence their respective microRNAs were found to be downregulated while the microRNAs for cldn5 were upregulted. This study proved hsa-miR-10a, hsa-miR-10b, hsa-miR-423, hsa-miR-497, hsa-miR-326, hsa-miR-424, hsa-miR-195 and hsa-miR-15a to be strongly associated with human ovarian cancer.

Chapter No 01

Introduction

1. INTRODUCTION

At the end of twentieth century the most exciting breakthrough happened in biological research when scientists discovered small RNA which then termed as microRNA. Further research suggested that they are involved in various fundamental cellular processes. Biological scientists were stunned by the discovery of hundred of genes encoding microRNA's which were found to be negative regulators of gene expression. That's why Meltzer (2005) stated them as "small RNAs with big impacts". The finding of involvement of microRNAs in cancer was a turning point in cancer research. Many studies have been done to elucidate the microRNA's involvement in cancer but due to the complexity of their function these researches still failed to show the complete picture of the problem.

1.1 MicroRNA

Recent advances in high-throughput technology have led to the most important discovery of the micro regulators of protein product, the MicroRNA. These are the endogenous RNAs of ~22nt length and they perform the regulatory roles by complementarily binding to the protein- coding genes and hence blocking their productive translation (Lai, 2003; Bartel, 2004). The output of protein products is immensely influenced by the critical role of microRNAs in regulation of coding genes. They mostly affect those genes which are involved in stem cell maintenance, developmental timing and other developmental and physiological processes in plants and animals (Carrington and Ambros, 2003). Brennecke *et al.* (2003) and Xu *et al.* (2003) explained the functional role of microRNAs in cell proliferation and fat metabolism in

flies. In plants the microRNAs are involved in control of leaf and flower development (Aukerman and Sakai, 2003; Chen, 2003; Emery *et al.*, 2003; Palatnik *et al.*, 2003).

The first microRNAs were discovered in *C. elegans*. These were let-7 and lin-4 which are the components of gene regulatory network which controls the larval developmental timings in *C. elegans* (Lee *et al.*, 1993; Wightman *et al.*, 1993; Moss *et al.*, 1997; Reinhart *et al.*, 2000; Abrahante *et al.*, 2003; Lin *et al.*, 2003). MicroRNAs are evolutionary conserved (Zhang *et al.*, 2006; Cai *et al.*, 2006; Lee *et al.*, 2007).

Other major qualities of microRNAs are their time, tissue (Liu *et al.*, 2009) and developmental stage specific (Malumbres *et al.*, 2008; Watanabe *et al.*, 2005) expression patterns. The relative expression of microRNAs in different tissues describes the more detailed insights about their functioning (Lagos-Quintana, 2002). Babak *et al.* (2004) describe that in adults microRNAs are mostly tissue specific. The microRNAs having similar expression are located close to each other (Liang *et al.*, 2007). These specifications provide the evidences for their role in the developmental cycles of the organisms.

The biological functioning of several microRNAs has been proved, for example miR-125b and let-7 role in cell proliferation, miR-181 in hematopoietic B-cell lineage, miR-430 in brain functioning, miR-143 for development of adipocyte and miR-15a and miR-16-1 in survival of B-cell (Harfe, 2005).

MicroRNA's are estimated to comprise between 0.5 and 1 percent of the predicted genes in worms, flies, and humans and hence they are most abundant class of regulatory genes in animals and it shows their extended functioning in regulation of

genes (Lagos-Quintana *et al.*, 2001; Lau *et al.*, 2001; Lee and Ambros, 2001; Lai *et al.*, 2003; Lim *et al.*, 2003a and Lim *et al.*, 2003b).

The prediction of involvement of two clustered miRNAs in human B-cell chronic lymphocytic leukemias (Calin *et al.*, 2002) made the scientists to ponder that how often microRNAs participates in human cancers. Further investigations proved that miRNAs were detected in almost every type of cancer including breast carcinoma, hepatocellular carcinoma (Murakami *et al.*, 2006), papillary thyroid carcinoma (He *et al.*, 2005), B-cell chronic lymphocytic leukemia (Calin *et al.*, 2004), lung cancer (Takamizawa *et al.*, 2004; Yanaihara *et al.*, 2006), primary glioblastoma (Ciafre, 2005), endocrine pancreatic tumors (Volinia *et al.*, 2006) and colorectal carcinoma (Micheal *et al.*, 2003). Now it is confirmed that microRNAs are involved in the pathogenesis of almost every cancer (Calin and Croce, 2006) because their involvement in fundamental pathways and their interaction with specific oncogenes are the undeniable facts.

The reasons for misexpression of miRNAs in cancers are not well understood but the basic point raised by scientists in this regard is their genomic location. They are normally found at breakpoint cluster regions or minimal regions of chromosomes. The latest study demonstrated that more than 50% of miRNA genes are located in cancer-associated genomic regions or in fragile sites (Calin *et al.*, 2004). Therefore miRNAs are deemed to play a crucial role in the initiation and progression of cancer in humans, and those with a role in cancer are designated as oncogenic miRNAs (oncomiRs) (Cho, 2007).

1.2 Ovarian Cancer

Ovarian cancer is the sixth most common cause of cancer death among woman worldwide and causes approximately 125,000 deaths annually (Cannistra, 2004). Only 30% of patients with ovarian cancer survive 5 years after initial diagnosis (Greenlee *et al.*, 2001). Ovarian cancer is a silent killer because of late stage diagnosis and very little survival rate. Environmental and genetic factors are both important in ovarian carcinogenesis. In the previous two decades, much progress has been made in identifying genes involved in the development of ovarian cancer. These known genes are useful in understanding the pathogenesis of ovarian cancer and defining its molecular signature. They can serve as biomarkers for early diagnosis and targets for drug development. The involvement of microRNAs in many cancer types is evident from research, so there are many chances that they might also be involved in ovarian cancer (Esquela-Kerscher and Slack, 2006). Several microRNAs had already found implicated in ovarian cancer such as up regulation of miR-200a and miR-141 and down regulation of miR-125b1 and miR-145 in ovarian cancer (Iorio *et al.*, 2007).

The complexity of disease and late stage diagnosis compel the scientists to unravel further molecular mechanisms involved in ovarian cancer so that early stage detection may become possible. The advancement in therapeutic research is required to reduce the mortality rate due to this most lethal gynecological malignancy.

Microarray analysis is the best high through-put techniques for microRNA genes involved in cancers (Kim and Nam, 2006; Liu *et al.*, 2004; Hammond, 2006). The current research is focused to extract most susceptible microRNAs in ovarian cancer. As there is no straight forward in silico method for digging microRNA genes so an indirect approach

was utilized by which first the highly expressive gene families in ovarian cancer were mined and then their respective microRNAs were predicted. The choice for selection of families was based upon their most expressiveness and also on novelty. This research is majorly narrowed down in to two gene families the pyruvate dehydrogenase kinase family and the claudin family.

1.3 Pyruvate dehydrogenase kinase family

Cancer is a metabolic disorder (Seyfried and Shelton, 2007). In cellular biology the word metabolic automatically refer to the organelle which, as being a power house of a cell, is considered as a supreme commander for all of the metabolic processes going on in human body, the Mitochondria (Stine,). In 1956, Otto Warburg was the first one to report the mitochondrial dysfunction in tumor cells (Warburg, 1956). His findings opened a new line of research and understanding for cancer. After Warburg many work was carried out regarding it (Cavalli and Liang 1998) and it was proved that tumor cell mitochondria has a disrupted energy metabolism rate as compared to the normal cell mitochondria (Mayevsky, 2009; Seyfried and Mukherjee, 2005; Chen *et al.*,2009; Ramanathan *et al.*, 2005). So it is necessary to unravel the molecular mechanisms in glycolysis and citric acid cycle which further involve in different cancers. In every cell the link of glycolysis with citric acid cycle is maintained by an irreversible process of oxidative decarboxylation of pyruvate to form acetyl coA. This reaction is catalyzed by pyruvate dehydrogenase complex which plays a crucial role in this reaction.

The glucose metabolism is regulated by pyruvate dehydrogenase kinase (PDK) which has the ability to switch off the mitochondrial pyruvate dehydrogenase complex (PDC). Four isoenzymes of pyruvate dehydrogenase kinase (PDK1, PDK2, PDK3 and

PDK4) with tissue specific activities have been identified in mammals so far (Bowker-Kinley *et al.*, 1998). The PDKs activity was mainly observed in case of fasting, starvation and in diabetes (Majer *et al.*, 1998; Guerre-Millo *et al.*, 2000). Increase in PDK4 activity in starvation and diabetes was shown in rat tissues by Wu *et al.* (1999). PDK2 is declared as a major isoenzyme responsible for regulation of pyruvate dehydrogenase complex (Gudi *et al.*, 1995). Although the major reported articles said that the contribution of these isoforms is mostly in muscles, heart and liver but due to critical role of these isozymes in energy metabolism, it is believed that any defect in their activity may become a cause for a cancer.

Tumor cells have increased energy demands and in order to fulfill it they try to up take more oxygen and nutrients then normal cells. This results in creation of hypoxic environment (Harris, 2002). Cancer cells undergo various adaptations in order to survive in unfriendly hypoxic environment. These adaptations are induced by Hypoxia-Inducible Factor (HIF) (Semenza, 2003; Gordan and Simon, 2007). HIF uses various different ways to regulate this mechanism and one of them is by switching off the conversion of pyruvate in to acetyl coA by blocking the activity of PDH complex through directly stopping PDK1 activity (Kim *et al.*, 2006). In the meanwhile estrogen- related receptors (ERR) stimulate increase in PDK4 expression in order to prevent aerobic metabolism of glucose (Wende *et al.*, 2005; Araki and Motojima, 2006; Zhang *et al.*, 2006). This mechanism allows the tumor cells to survive in hypoxic environment and fulfill the high energy demand of these cells (Ao *et al.*, 2008). The crucial role of PDKs in survival of tumors is well understood by this mechanism. The increased expressions of PDK1 in

breast cancer (Maurer *et al.*, 2009; Lin *et al.*, 2005) and decrease of PDK4 in cervical cancer (Carlson *et al.*, 2007) also validate presence of their pivotal role in cancers.

1.4 Claudin gene family

In multicellular organisms, there are epithelial cellular sheets which function as diffusion barriers, to maintain distinct fluid compartments and also they are involved in active transport of materials across the barrier, to dynamically establish environment of each inter-compartmental unit. These physiological functions demand a seal to the diffusion of solutes on cellular sheets. Tight junctions (TJ) are considered responsible for this intercellular sealing (Anderson *et al.*, 1995; Goodenough, 1999; Tsukita *et al.*, 2001). These tight junctions are not simple barriers because they control selectivity of ion and size. Their functioning also alters according to type of cell and various other physiological requirements. This permeability criterion is essential for maintaining dynamic environmental conditions of each compartment (Powell, 1981; Reuss, 1963). Tight junctions are formed by tight junction strands surrounded by integral membrane proteins and transmembrane domains and labeled as the claudins (Tsukita, 2008).

The claudin (*CLDN*) gene family is responsible for encoding claudin proteins which plays role in formation of tight junctions (Oliveria, 2006). Claudins are transmembrane proteins that extend the bilayer at the spot where N and C terminal are towards the cytoplasmic region (Findley, 2009). Claudins are called paracellular barriers because they controls paracellular permeability across the epithelial cells and due to this they have role in establishment and maintenance of apico-basal polarity (Shin *et al.*, 2006). They form the seal and they have pore-like properties. This family consists of 23

tight junction proteins (Sukita and Furuse, 2000). The correct arrangement of all claudin genes is very necessary to perform its function which is the formation of tight junctions. Any problem in its gene arrangement causes cancers. Claudins assist the permeability of endothelial and epithelial cells and that's why the reduction in tight junctions is a key step towards cancer progression (Martin, 2001). Claudin gene family is implicated with various types of cancers (Morin, 2000; Swisshelm *et al.*, 2005; Nakanishi *et al.*, 2008; Dhawan *et al.*, 2011). Association of ovarian cancer with some members of the claudin family has already been reported before e.g. cldn3 (Heinzelmann-Schwarz *et al.*, 2004), cldn4 (Boylan *et al.*, 2011) and cldn7 (Hough *et al.*, 2000; Rangel *et al.*, 2003; Hewitt *et al.*, 2006; Bignotti *et al.*, 2006). The function of claudins is highly tissue specific because claudin 3 and claudin 4 was observed in ovarian cancer but not in ovarian cystadenomas (Rangel *et al.*, 2003).

In present research the pathogenesis of PDK and CLDN gene family was estimated first and then their respective microRNAs were predicted. This methodology will finally provide the most susceptible gene targets and their respective microRNAs for human ovarian cancer and also provide a novel paradigm for ovarian cancer therapeutic research by representing the full picture of gene expression.

Chapter No 02

Materials and Methods

2. MATERIALS AND METHODS

The methodology includes two simple steps. First, to identify the genes associated with human ovarian cancer and secondly to find the respective microRNA's for those genes. In order to complete the first task two ovarian data sets were downloaded from Gene Expression Omnibus (GEO). High throughput data is publicly available on this repository. As the data is generated from microarray and next generation sequencing methods that's why the data on GEO is very reliable because of its actual collection of data (Edgar *et al.*, 2002) (Barret *et al.*, 2005; Barrett and Edgar, 2006; Barret *et al.*, 2007).

These two datasets GSE4122 and GSE6008 contain expression amounts for both normal and malignant ovaries.GSE6008 contain 99 individual of ovarian tumors and 4 individual of normal ovary samples contributed by Hendrix and the second dataset GSE4122 was contributed by Tate and his co workers with 32 cancerous tumors and 14 controls. These datasets were mined for finding novel gene markers in human ovarian cancer by systematic investigation. This was done by comparing the normal ovarian expression data with the cancerous one. The assessment was made by employing various statistical and clustering methods. This evaluation was performed by utilizing the techniques in a combination and series which finally divided most vulnerable targets for ovarian cancer.

For the microarray data analysis the major software used was Multi Experiment Viewer (Mev) version 4.0 (Saeed *et al.*, 2003; Saeed *et al.*, 2006). This is a part of TM4 software suite for all type of microarray analysis. It is comprised of statistical, clustering

and classification methods. It is the best known data mining software for transcriptome analysis. Only few statistical and clustering algorithms were used in this research from this tool including t-test, Significance Analysis of Microarrays (SAM), Hierarchical clustering, K-means clustering and Self organizing tree algorithm (SOTA). MicroRNA prediction is the second phase of study. For assigning microRNA's various prediction tools were applied. The methodology for both phases of investigation is explained below.

2.1 Mining susceptible genes for ovarian cancer

For finding the most suitable target genes in ovarian cancer a variety of statistical and clustering algorithms were used. Along with them two tools were subjected for verification of microarray analysis results.

2.1.1 Statistical methods

The huge size and the implicit variant nature of microarray data makes it very difficult to extort the hidden biological information from microarray data. In order to deal with this bulk of data statistical techniques are used. Two simple and important statistical tools utilized in this research were t-test and Significance analysis of microarrays (SAM).

T-test is the simplest known technique for comparison of two phenotypes in microarray data sets mostly these two conditions is normal and diseased. The result of this comparison is in the form of p-value or the probability of expression of each single gene. The expression values for each gene from all experiments are taken to the account for assessing the differential expression of each gene. As the two downloaded sets contain data of two phenotypes so the t-test performed was "Between paired subjects" t-

test. From the data sets only those genes which have probability less than 0.005 ($P<0.005$) were considered significant. The output of this test not only provides p-values for each gene based on t-distribution but it also calculates the false discovery rate (FDR) which shows the expected false positives. The statistical significance of a gene is dependent upon both p-value and FDR. The t-test results were displayed on volcano plot in order to visually explain the differential expression of genes. The scatter plot best shows the behavior of the genes. It is the graph between negative $\log_{10}$-transformed *p*-values and difference of mean of expression between both phenotypes. The purpose of doing t-test was to extract the most significant genes from huge microarray data sets.

Although t-test is the very reliable tool for statistical analysis but it is a very basic test and some times its result were biased only on p-value while the fold change was very small so its enhanced version was also used in order to exclude the biased results (Cui and Churchill, 2003). This enhanced statistical test is Significance Analysis of Microarrays (SAM). It is also called as the S-test (Tusher, 2001). In SAM the genes with greater fold change and significant p-values were only considered as significant. The cut-off value was set to >2 so that only those genes which have fold change greater than two between two phenotypes were selected. The SAM results can be clearly understood by SAM graph which separates the positive and negative significant genes on the basis of observed and expected value. The genes common in both t-test and SAM were selected for further analysis.

2.1.2 Clustering

Cluster analysis was performed on the similar genes extracted from t-test and SAM. The purpose of performing clustering after statistical testing is to reduce the chances of error as it was not necessary that the expression of statistically significant gene was also significant when there was a matter of comparison between the normal and cancerous gene data. Clustering is a very powerful technique for analyzing gene expression data (Jiang *et al.*, 2004). It is an unsupervised learning process by which similar genes are grouped together. Cluster analysis reveals the correlated pattern of gene expression (Michaels *et al.*, 1998). In order to find inter related genes from the significant set of genes three different clustering techniques were applied. Although there are many clustering techniques but in this research only three techniques were applied namely hierarchical clustering, K-means clustering and Self Organizing Tree Algorithm (SOTA).

Hierarchical clustering is the easiest and the most important basic level technique of clustering. The gene tree is formed by agglomeratively or divisively grouping the clusters of genes having similar characteristics. All of the selected significant genes were gone through the hierarchical clustering algorithm and as a result the microarray gene pattern showed regions with significant difference in the normal and ovarian samples of genes. Those regions with significant expression patterns were then specified for further investigation.

K-means clustering is the second clustering method utilized in order to exclude those false negatives (if any) which were accidently included in the hierarchical cluster as most significant hit. In this technique user defines the number of clusters that's why it is

known as k-means. The clusters which represent more differential expression visually were collected for further research.

Self organizing tree algorithm (SOTA) is the more superior and efficient clustering algorithm as it uses neural networks for clustering. It works in a divisive manner and efficiently adds cluster by utilizing top to bottom approach (Herrero *et al.*, 2000). It generates dendogram which summarizes the result for each cluster. Each row is the representative of a cluster. By only visualizing the dendogram the expression of each cluster can be judged. The most significant clusters were then picked by analyzing dendogram pattern. Pearson correlation distance matrix was selected because it is best matrix for analyzing the gene expression patterns (Haeseleer, 2005). The average linkage clustering parameter was chosen for analysis because in it the cluster differences were measured by taking average of all data point distances in a cluster. Hence SOTA gave the more insights in the expression patterns of the genes.

The genes having significant expression profiles shown in all the three clustering methods were then listed down for further analysis. The genes which fail to appear significant in any of these tests were excluded from the list of significant genes chosen for analysis by the statistical tests.

The genes which passed both the statistical and clustering tests were then reviewed in literature to know that if they were already predicted in association with ovarian cancer. Many of the extracted genes were already predicted but here some novel targets were also observed. Fortunately they belong to the same families and by this the representation of two families in ovarian cancer was identified. Those two pyruvate

dehydrogenase kinase (PDK) and claudin families were further verified for their association with ovarian cancer and then their potential microRNA's were predicted in the second phase of research.

2.1.3 Serial Analysis of Gene Expression (SAGE) analysis

To perform serial analysis of gene expression in silico the web tool SAGE Genie (Liang, 2002) was employed, available at http://cgap.nci.nih.gov/SAGE/AnatomicViewer. SAGE Genie is a part of Cancer Genome Anatomy Project (CGAP). It is a very efficient tool for transcriptome analysis. It contains the mRNA libraries of both cancer and normal tissues. The expression level of above separated genes for further investigation was calculated by counting tags per million (TPM) from the available expression libraries. Only the reliable tags were picked and then mapped with their unigene clusters. The TPM values with fold differences between cancerous and normal ovarian libraries greater than two fold was considered significant. The reason for doing SAGE analysis is to validate the microarray results.

2.1.4 GENT: Gene Expression Database of Normal and Tumor Tissues

One of the verification tools used for the microarray results was GENT (Shin *et al.*, 2011). It is basically a database which contains both normal and cancerous tissue data for comparison of expression pattern, and contains updated information about gene expression. The result from microarray and SAGE were again confirmed from GENT which graphically shows the levels of expression of those genes in different tissues. From here the expression level of the genes were only observed in normal and malignant ovaries.

2.2 Predicting MicroRNA targets

The genes having novel association with ovarian cancer obtained after mining of ovarian cancer data sets were then checked for their corresponding microRNA's. Five target prediction tools were used for prediction as shown in table 2.1. These were targetscan, miRanda, pictar, miRGen and RNAHybrid. TargetScan is a very efficient and reliable tool for prediction of microRNA targets. It works by locating the conserved pairing 5' region of miRNA with the mRNA. Then the identified hits are evaluated and quality of output is checked.

miRanda is an algorithm based upon dynamic programming. It is the extension of smith waterman algorithm with the addition of position dependent weights. Sequence conservation was also taken in to account while predicting the miRNA targets for mRNA. The microrna.org is the resource for miRNA target prediction and predictions were based upon miRanda algorithm. Pictar predicts microRNA by literature mining so it is the most reliable one but at the same time it is very limited in the sense that very miRNA's research to date is not sufficient. miRGen is basically a database providing miRNA data from 11 animal genomes. It is user friendly and reliable repository for miRNA's. RNAHybrid predict microRNA by calculating the binding energies. It is a tool which finds the minimum free energy hybridization of mRNA of target gene and microRNA. The tight binding accounts for the most appropriate functioning of microRNA's. Those microRNA that bind more correctly with genes than others were more likely to be involved in the regulation of the respective gene.

Table 2.1 Tools used for microRNA target prediction.

Tool name	URL	References
TargetScan	http://www.targetscan.org/	Lewis *et al.* (2005) Grimson *et al.* (2007) Friedman *et al.* (2009)
miRanda	http://www.microrna.org/microrna/	Enright *et al.* (2003) John *et al.* (2004) Batel *et al.* (2010, 2008)
Pictar	http://pictar.mdc-berlin.de/	Grun *et al.* (2005) Krek *et al.* (2005) Lall *et al.* (2006) Chen & Rajewsky (2006)
miRGen	http://www.diana.pcbi.upenn.edu/miRGen.html	Megraw *et al.* (2006)
RNAhybrid	http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/	Rehmsmeier *et al.* (2004) Krüger *et al.* (2006)

Chapter No 03

Results

3. RESULTS

The scanning of high throughput ovarian data resulted in a number of differentially expressed genes. The process of scrutinization was highly sensitive and thus come up with the various novel results. The sequence of statistical and clustering checks arises with collection of noteworthy gene targets in ovarian cancer. Many of the genes predicted through current research were already reported but there were also few genes whose association with cancer of ovaries was never explored before. The shortlisted candidate genes for ovarian cancer are shown in table 3.1.

From all these shortlisted genes only few genes were selected for further investigation. After looking in to the gene list in table 3.1 it was revealed that some genes belongs to same families e.g., cldn3, cldn4, cldn7 belongs to the family of claudins. Similarly pdk2 and pdk4 represent the family of pyruvate dehydrogenase kinase and in the same way slc25A17, slc4A3, slc18A2 and slc16A4 are members of solute carrier family. Claudin and PDK gene family were only picked for further analysis because these two families showed greater representation with significant expression patterns. The representation of two or more genes from a same family shows the involvement of whole gene family in ovarian cancer hence it was necessary to investigate those gene families rather than separate genes. It does not mean that individual genes were not significant but there is a possibility that there appearance is by chance or might be they are false negatives. So at this stage in order to limit the search to the hottest genes these two families were chosen.

Table 3.1 The candidate genes for ovarian cancer.

Reference ID	Gene symbol	Gene Name
203453_at	SCNN1A	sodium channel, nonvoltage-gated 1 alpha
201596_x_at	KRT18	keratin 18
202005_at	ST14	suppression of tumorigenicity 14
202826_at	SPINT1	serine protease inhibitor, Kunitz type 1
214352_s_at	KRAS	Kirsten rat sarcoma 2 viral oncogene
213693_s_at	MUC 1	mucin 1
201428_at	CLDN4	Ciaudin 4
202790_at	CLDN7	Claudin 7
209008_x_at	KRT8	keratin 8
212070_at	GPR56	G protein-coupled receptor 56
203953_s_at	CLDN3	Claudin 3
202525_at	PRSS8	protease, serine, 8 (prostasin)
210827_s_at	ESE-1	epithelial-specific transcription factor
203892_at	HE4	Homo sapiens epididymis-specific
201839_s_at	TACSTD1	tumor-associated calcium signal transducer 1
200606_at	DSP	desmoplakin
209335_at	DCN	decorin
219778_at	FOG2	Friend of GATA2
209200_at	MEF2C	myocyte enhancer factor 2C
205792_at	WISP2	WNT1 inducible signaling pathway protein 2

Reference ID	Gene symbol	Gene Name
219935_at	ADAMTS5	aggrecanase-2
205082_s_at	hAO	hydroxyacid oxidase (glycolate oxidase)
205381_at	P37NB	leucine-rich repeat (LRR)
204464_s_at	EDNRA	endothelin receptor type A
205051_s_at	KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
216733_s_at	GATM	glycine amidinotransferase
205158_at	RNASE4	ribonuclease
202747_s_at	ITM2A	integral membrane protein 2A
204719_at	ABCA8	ATP-binding cassette, sub-family A (ABC1), member 8
205960_at	PDK4	Pyruvate dehydrogenase kinase enzyme 4
219985_at	HS3ST3A1	heparan sulfate (glucosamine) 3-O-sulfotransferase 3A1
204260_at	CHGB	chromogranin B
208209_s_at	C4BPB	complement component 4-binding protein
214210_at	SLC25A17	solute carrier family 25 , member 17
210001_s_at	SOCS1	suppressor of cytokine signaling 1
33767_at	NEFH	neurofilament, heavy polypeptide
205981_s_at	ING1L	inhibitor of growth family, member 1-like
204967_at	APXL	apical protein, Xenopus laevis-like
206812_at	ADRB3	adrenergic, beta-3-, receptor

Reference ID	Gene symbol	Gene Name
209869_at	ADRA2A	adrenergic, alpha-2A-, receptor
205517_at	GATA4	GATA binding protein 4
207864_at	SCN6A	sodium channel, voltage-gated, type VI
203859_s_at	PALM	paralemmin
219914_at	ECEL1	endothelin converting enzyme-like 1
204381_at	LRP3	low density lipoprotein receptor-related
203058_s_at	PAPSS2	3'-phosphoadenosine 5'-phosphosulfate synthase 2
201556_s_at	VAMP2	vesicle-associated membrane protein 2
206879_s_at	NRG2	neuregulin 2
204287_at	SYNGR1	synaptogyrin 1
219414_at	CS2	calsyntenin-2
205918_at	SLC4A3	solute carrier family 4, anion exchanger, member 3
221215_s_at	ANKRD3	ankyrin repeat domain 3
205453_at	HOXB2	homeobox B2
205404_at	HSD11B1	hydroxysteroid (11-beta) dehydrogenase 1
205208_at	FTHFD	formyltetrahydrofolate dehydrogenase
48031_r_at	C5orf4	chromosome 5 open reading frame 4
214279_s_at	NDRG2	N-myc downstream-regulated gene 2
206756_at	CHST7	carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 7

Reference ID	Gene symbol	Gene Name
203426_s_at	IGFBP5	insulin-like growth factor binding protein 5
210751_s_at	SMP-30	regucalcin (senescence marker protein-30)
205116_at	LAMA2	laminin, alpha 2
203950_s_at	CLCN6	chloride channel 6
205079_s_at	MPDZ	multiple PDZ domain protein
219370_at	REPRIMO	candidate mediator of the p53-dependent G2 arrest
213030_s_at	PLXNA2	plexin A2
207233_s_at	MITF	microphthalmia-associated transcription factor
221234_s_at	BACH2	basic leucine zipper transcription factor 2
202590_s_at	PDK2	Pyruvate dehydrogenase kinase 2
205486_at	TESK2	testis-specific kinase 2
218486_at	KLF11	Kruppel-like factor 11
220332_AT	CLDN16	Claudin 16
206286_s_at	TDGF1	teratocarcinoma-derived growth factor 1
204154_at	CDO1	cysteine dioxygenase, type 1
203939_at	NT5	Homo sapiens 5 nucleotidase
204482_AT	CLDN5	Claudin 5
205498_at	GHR	growth hormone receptor
203603_s_at	ZFHX1B	zinc finger homeobox 1B
204396_s_at	GPRK5	G protein-coupled receptor kinase 5
215506_s_at	DIRAS3	GTP-binding RAS-like 3

Reference ID	Gene symbol	Gene Name
203562_at	FEZ1	fasciculation and elongation protein zeta 1
213176_s_at	LTBP4	latent transforming growth factor beta binding protein 4
205845_at	CACNA1H	calcium channel, voltage-dependent, T type, alpha 1H
205803_s_at	TRPC1	transient receptor potential cation channel, subfamily C, member 1
203501_at	PGCP	plasma glutamate carboxypeptidase
204753_s_at	HLF	hepatic leukemia factor
218878_s_at	SIRT1	Sirtuin 1
220992_s_at	C1ORF25	N2-dimethylguanosine tRNA methyltransferase
213415_at	CLIC2	chloride intracellular channel 2
203787_at	SSBP2	single-stranded DNA binding protein 2
204201_s_at	PTPN13	protein tyrosine phosphatase, non-receptor type 13
203845_at	KAT2B	K(lysine) acetyltransferase 2B
205934_at	PLCE	phospholipase C, epsilon
205234_at	SLC16A4	solute carrier family 16 (monocarboxylic acid transporters), member 4

3.1 Pyruvate Dehydrogenase Kinase Family

The statistical analysis of microarray datasets GSE4122 and GSE6008 of ovarian cancer exposed significant genes highlighted in the expression pattern. The t-test and SAM hauled the most reliable genes differentially expressed in the malignant and normal ovarian data sets. Many of the extracted genes association with ovarian cancer were already reported. PDK2 and PDK4 were among with those extracted genes. These two members of PDK gene family was never identified previously for their connection with human ovarian cancer.

Up regulation of PDK2 and PDK4 in t-test was observed in both datasets, while in SAM of GSE6008 only PDK4 expression was seen as significant gene while SAM declared pdk2 as non significant gene because of its low fold change. The results of statistical tests were shown in the form of volcano plots figure 3.1 and figure 3.2 and SAM graphs figure 3.3 and figure 3.4. The significance of these genes was confirmed by their p-value and fold changes obtained from these tests (Table 3.2).

Clustering analysis also revealed almost the same results. Figure 3.5 and figure 3.7 summarizes the clustering results of PDK2 from GSE6008 and GSE4122 data sets respectively. The Hierarchical clustering, K-means clustering and clusters originated through self organizing tree algorithm (SOTA) having selected genes were displayed in these figures. Similarly figure 3.6 and 3.8 represents clusters of PDK4 from GSE6008 and GSE4122 respectively. Verification of PDK2 and PDK4 over expression in human ovarian cancer was done through SAGE where both tissue and cell line libraries of normal and cancerous ovaries were scanned. In cell lines normally there was no expression of PDK2 and PDK4. PDK2 over expression verified in cancerous tissues and cell lines but PDK4 expression result was

very different as it is up regulated in case of cell lines but it shows no or little expression in tissues (Table 3.3). Expression of PDK2 and PDK4 was also checked by GeneFinder tool (Figure 3.9 and Figure 3.10) which also verified SAGE results. Similarly when pdk4 was analyzed using GENT it was noted that in U133Plus2 it was up regulated (Figure 3.11a and Figure 3.11b), while in data set U133A its expression is down regulated (Figure 3.11c and Figure 3.11d). PDK2 over expression was also verified by both data sets of GENT (Figure 3.12)

Targeting microRNA's for PDK genes were predicted by using various tools and common results from these software were taken only. The binding energies were also calculated and finally only those microRNA's were shortlisted which were common output of all prediction tools and whose binding energies were also considered minimum (Table 3.4).

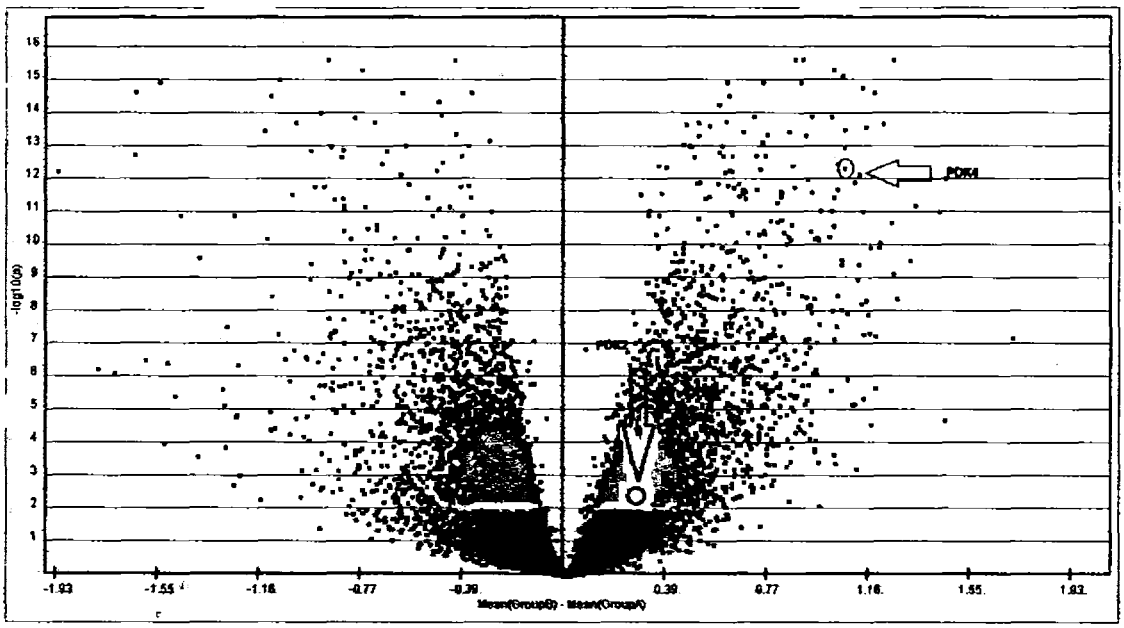


Figure 3.1 Volcano plot of PDK2 and PDK4 (GSE6008)

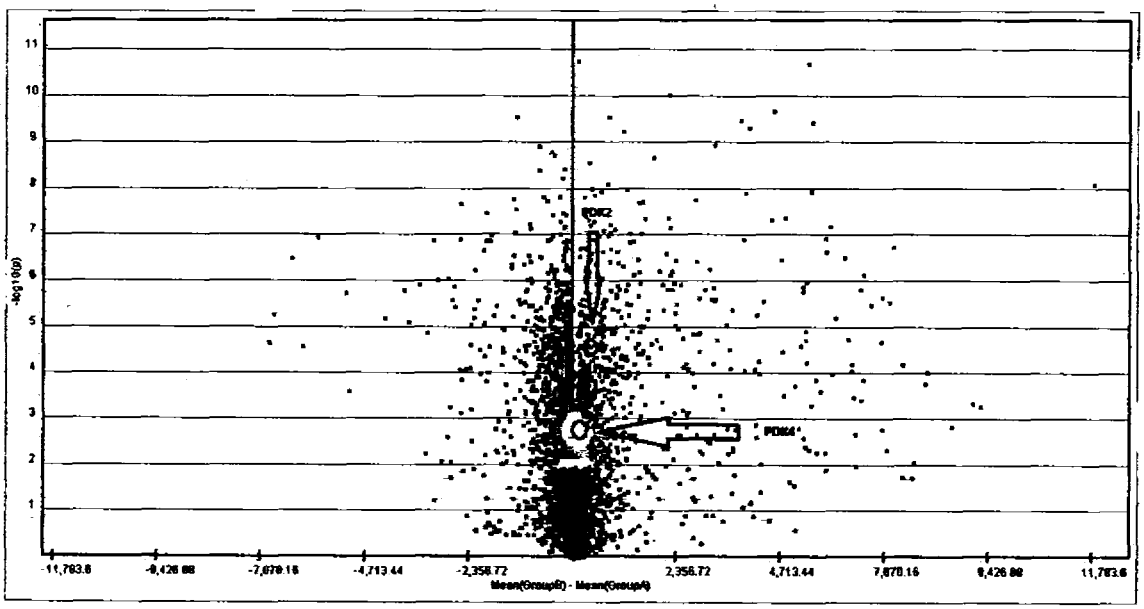


Figure 3.2: Volcano plot of PDK4 and PDK2 (GSE4122)

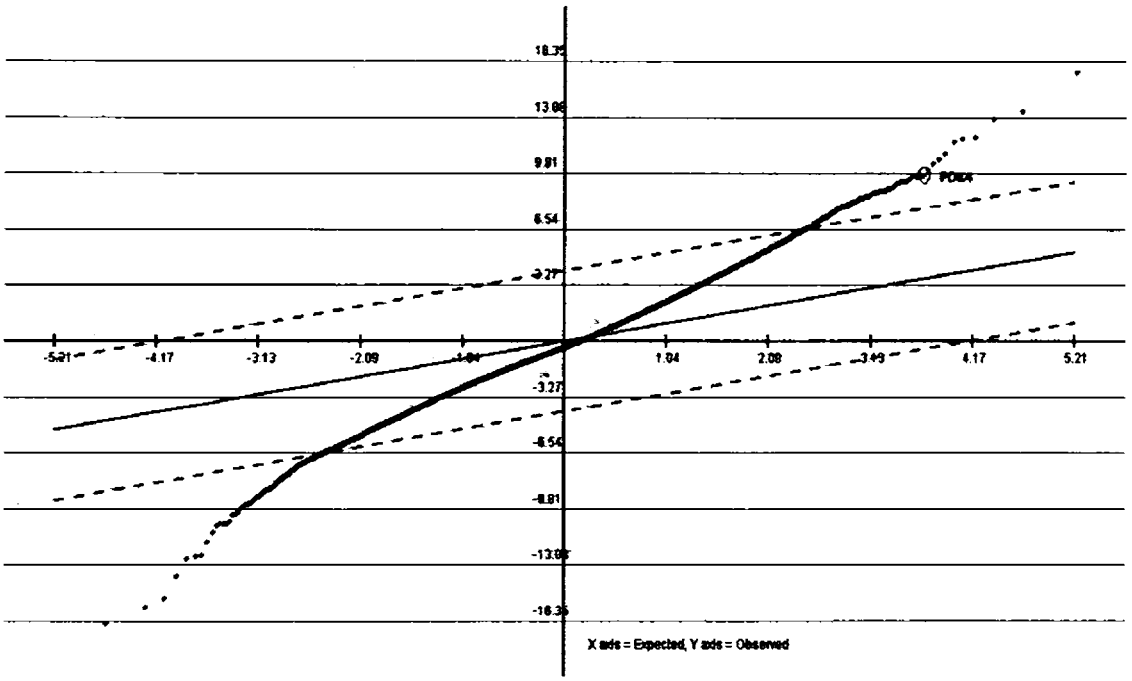


Figure 3.3: SAM graph of PDK4 (GSE6008)

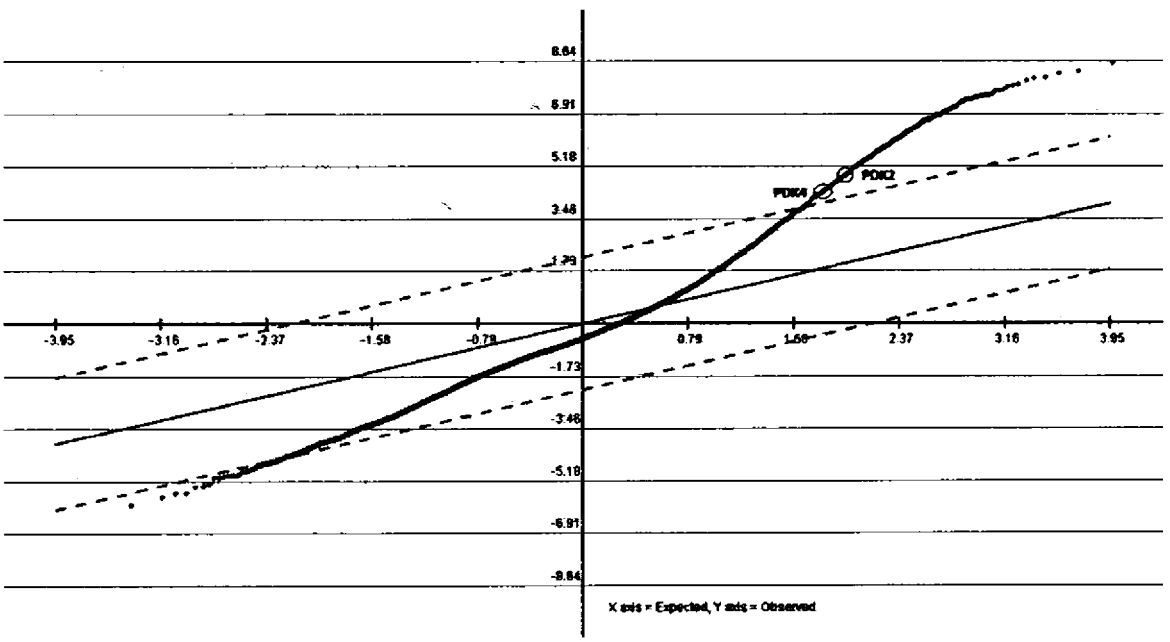
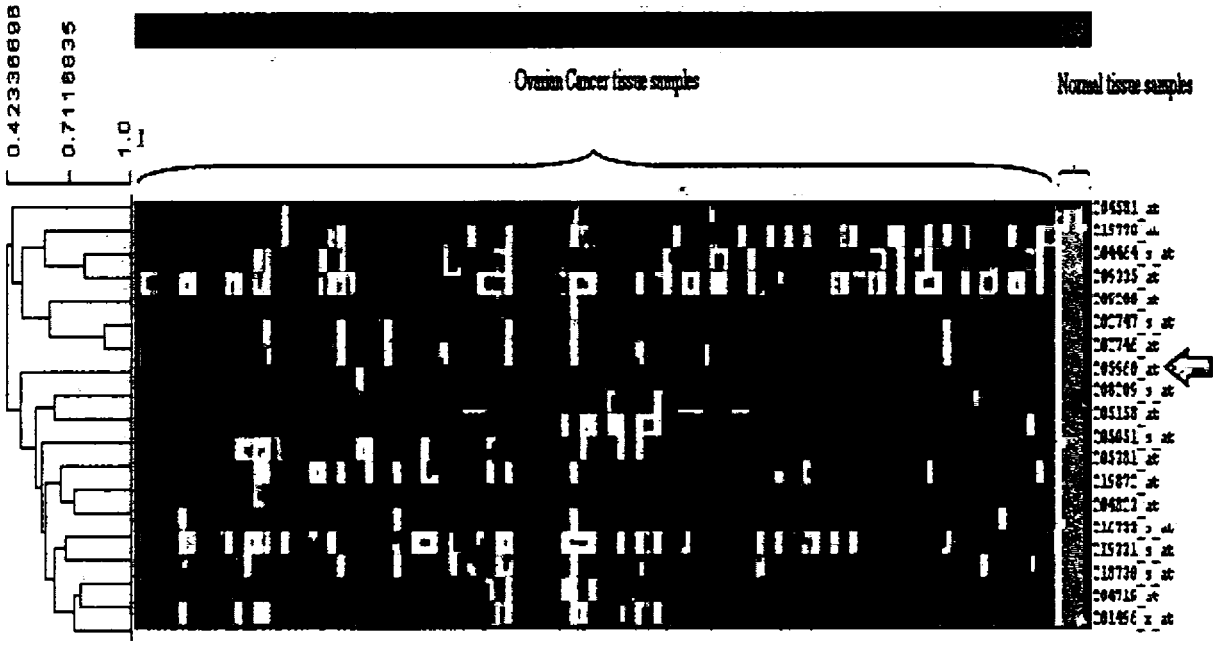
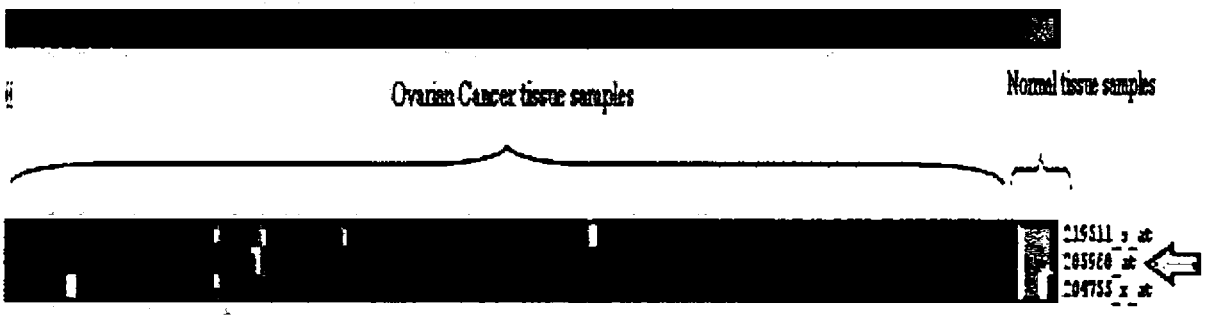


Figure 3.4: SAM graph of PDK4 and PDK2 (GSE4122)

(a)


0.42336698
0.71116835
1.0
Ovarian Cancer tissue samples
Normal tissue samples

(b)


Ovarian Cancer tissue samples
Normal tissue samples

(c)

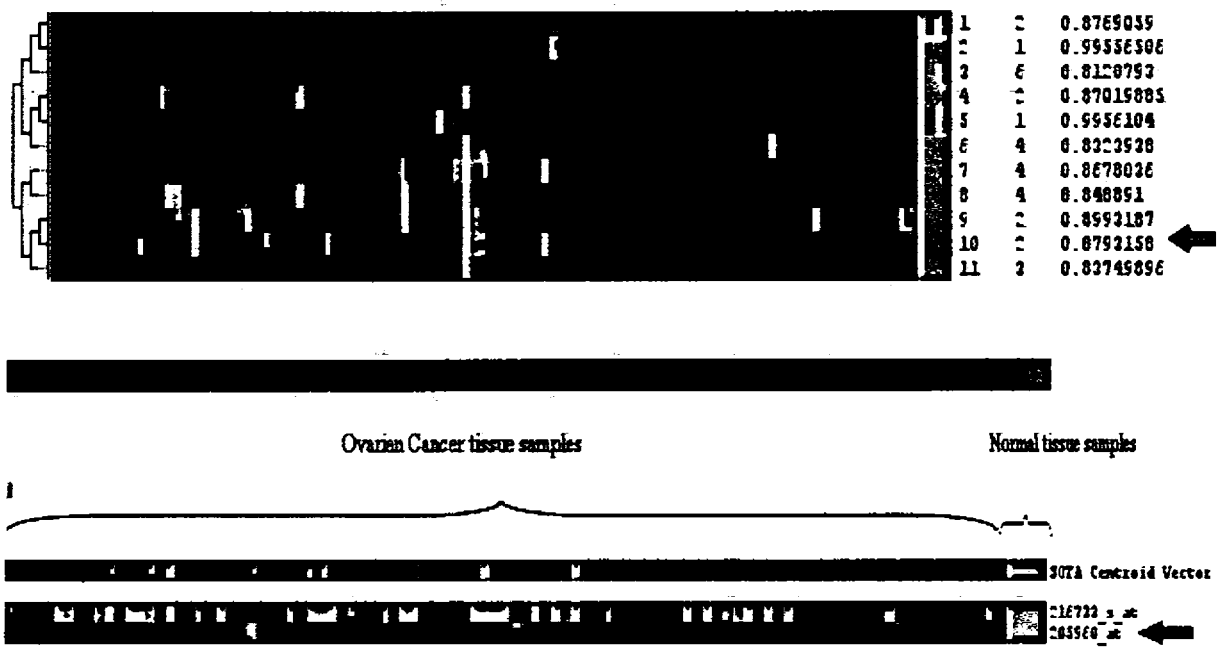
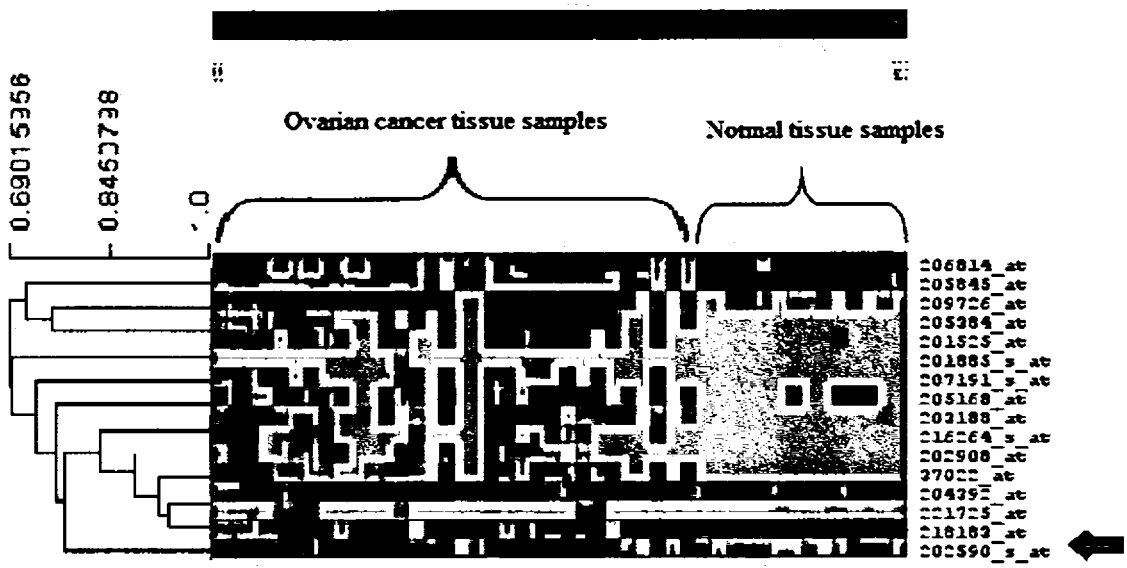
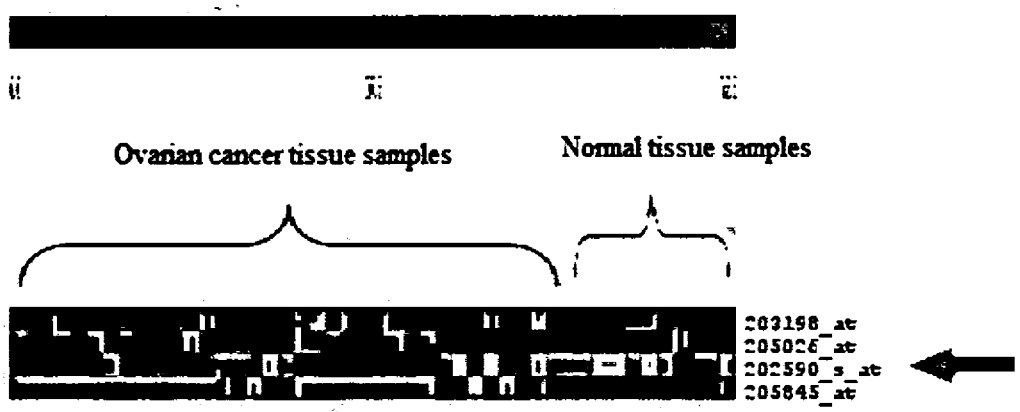


Figure 3.6: Clustering diagrams of PDK4 from data set GSE6008 (a) Hierarchical cluster (b) K-means cluster (c) SOTA cluster

(a)


0.6901536
0.8453738
1.0
Ovarian cancer tissue samples
Normal tissue samples
206814_at
205845_at
209726_at
205384_at
201525_at
201885_s_at
207191_s_at
205168_at
202188_at
216264_s_at
202908_at
37022_at
204392_at
221725_at
218183_at
202590_s_at

(b)


Ovarian cancer tissue samples
Normal tissue samples
203196_at
205026_at
202590_s_at
205845_at

(c)

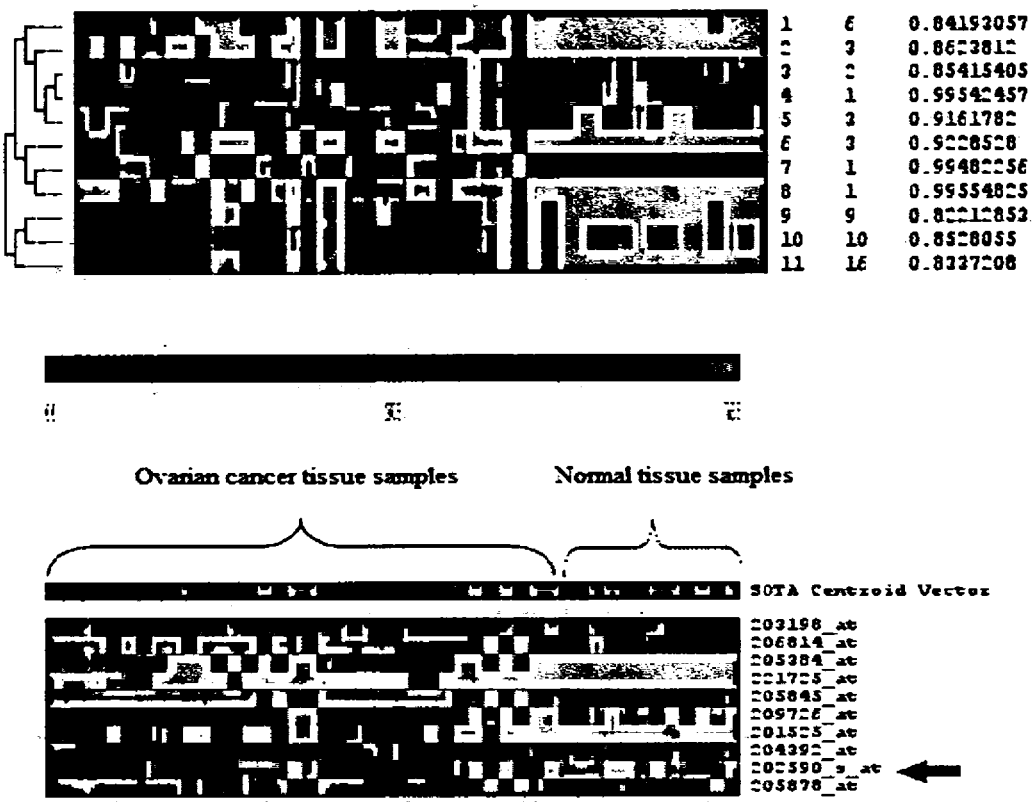
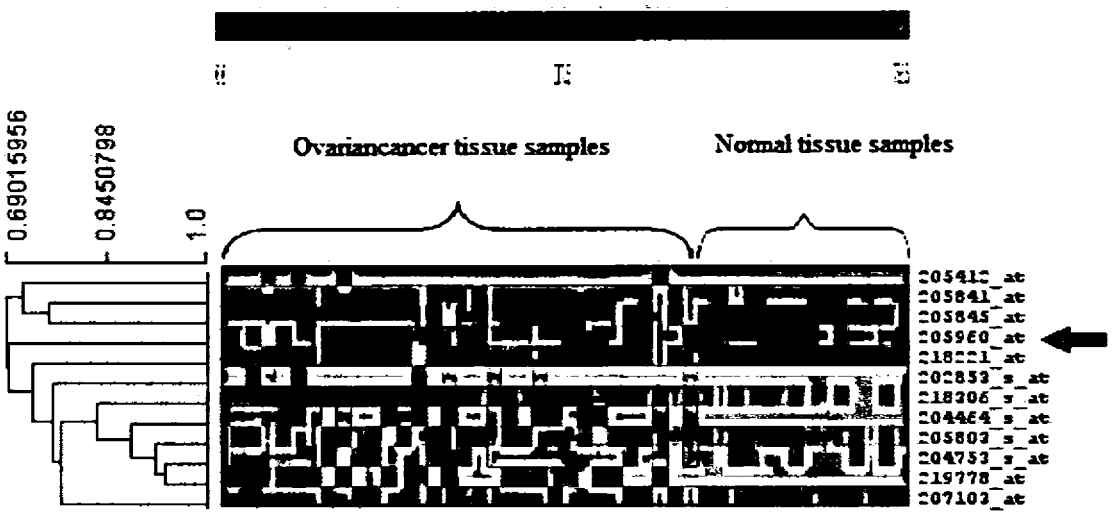
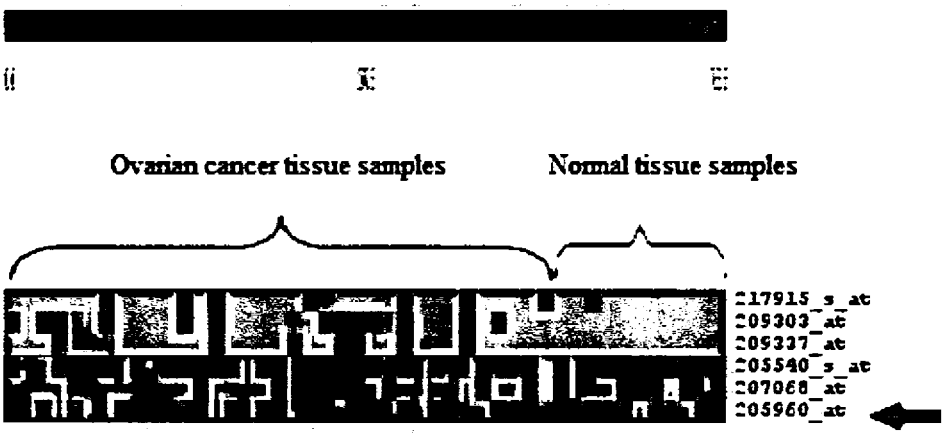


Figure 3.7: Clustering diagrams of PDK2 from data set GSE4122 (a) Hierarchical cluster (b) K-means cluster (c) SOTA dendogram and cluster

(a)


0.69015956
0.8450798
1.0
Ovariancancer tissue samples
Normal tissue samples
205960_at
217915_s_at
209303_at
209337_at
205540_s_at
207068_at
205960_at

(b)


Ovarian cancer tissue samples
Normal tissue samples
217915_s_at
209303_at
209337_at
205540_s_at
207068_at
205960_at

(c)

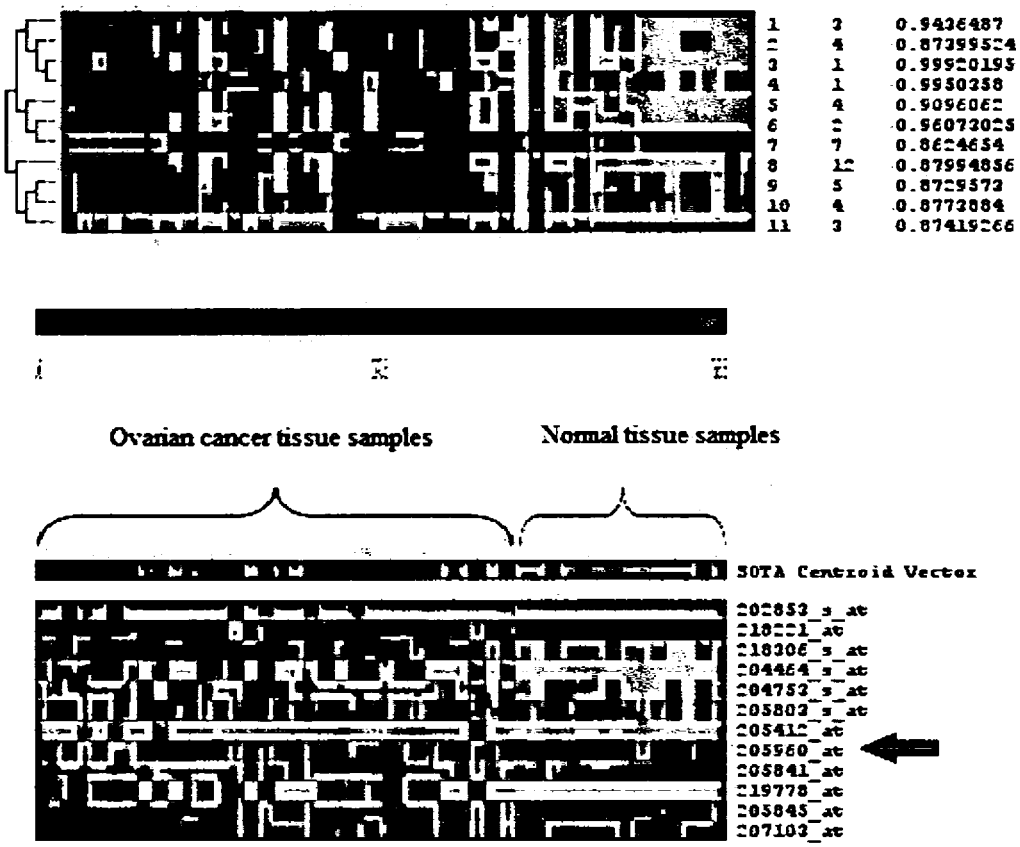


Figure 3.8: Clustering diagrams of PDK4 from data set GSE4122 (a) Hierarchical cluster (b) K-means cluster (c) SOTA dendogram and cluster

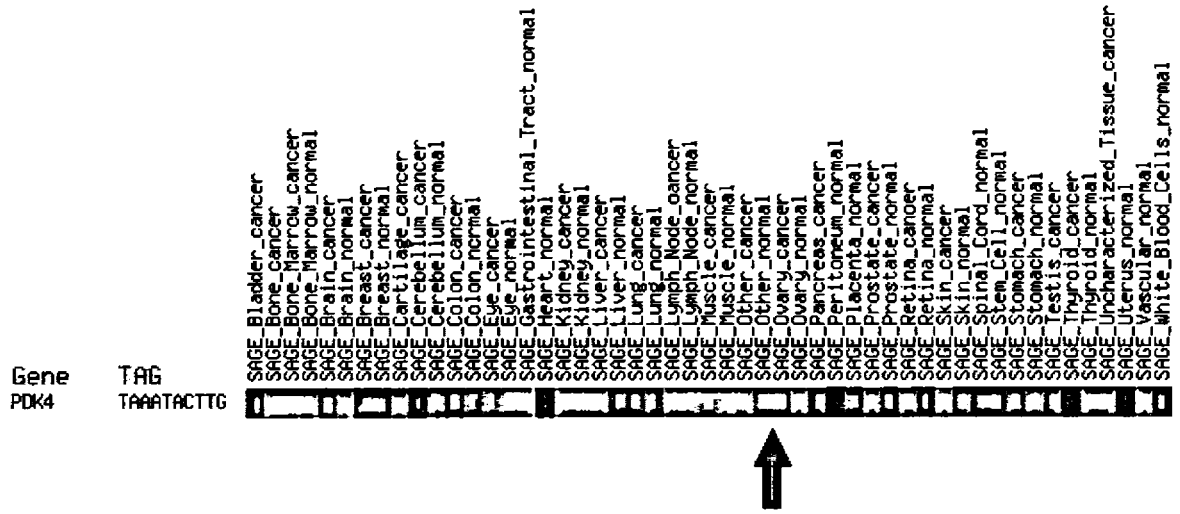


Figure 3.9: Short SAGE map of PDK4 from Gene Finder tool showing reduced expression in cancerous ovary

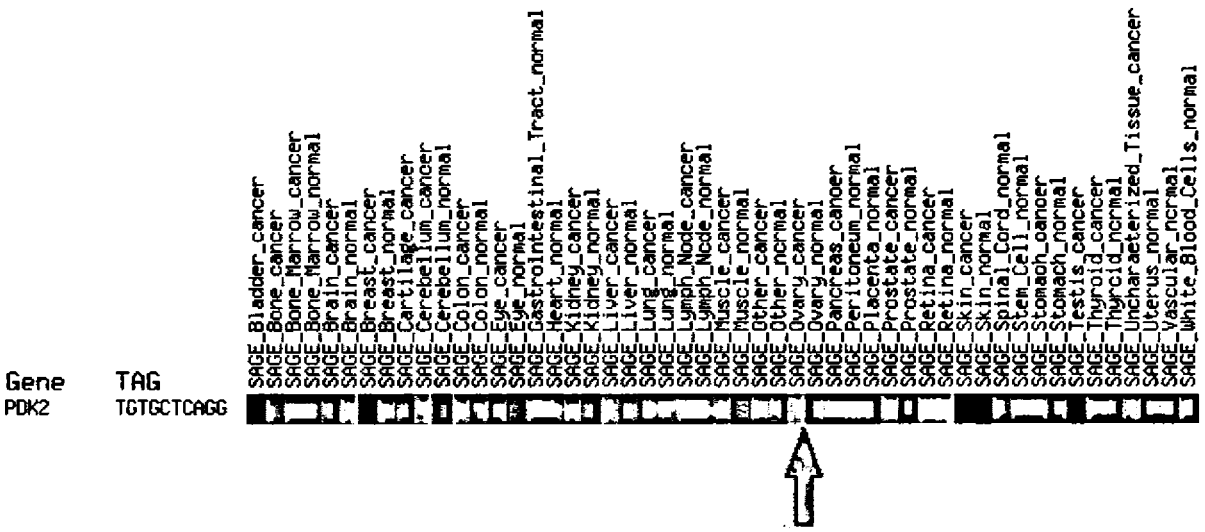
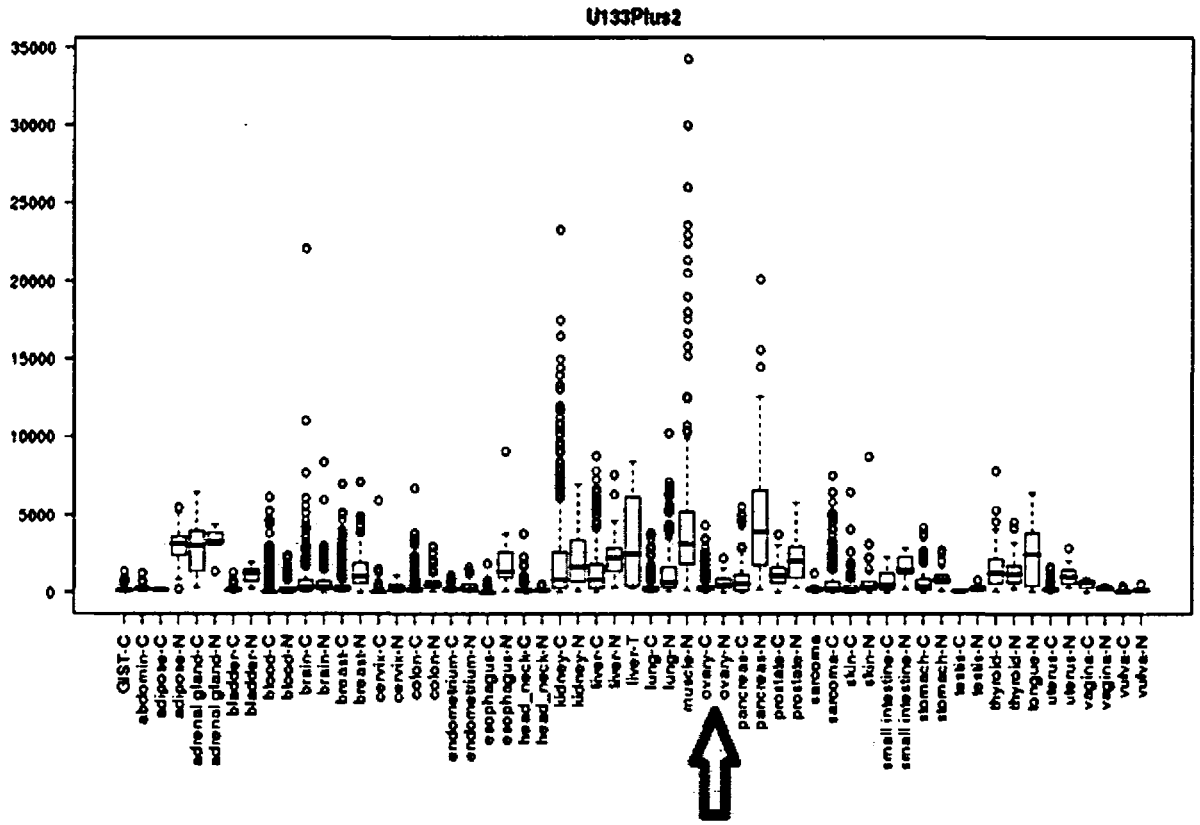
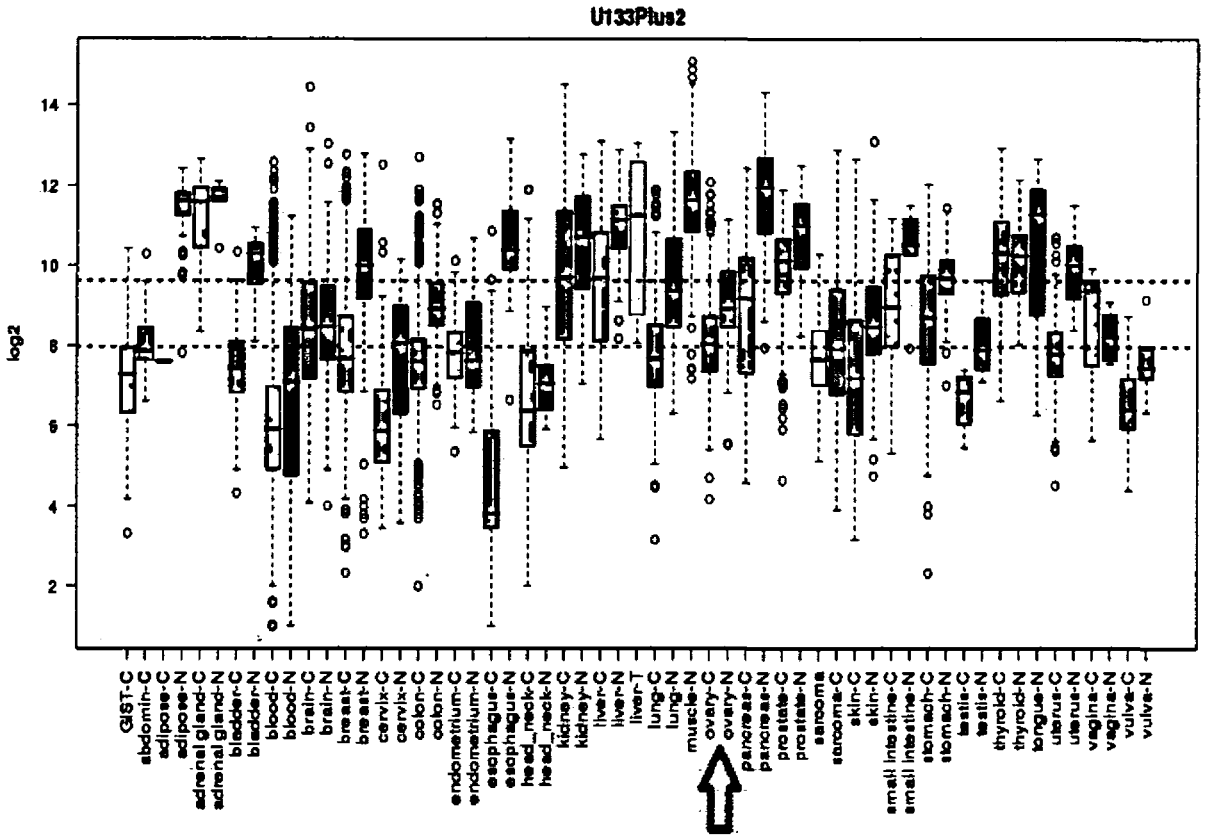


Figure 3.10: Short SAGE map of PDK2 from Gene Finder tool showing over expression in cancerous ovary.

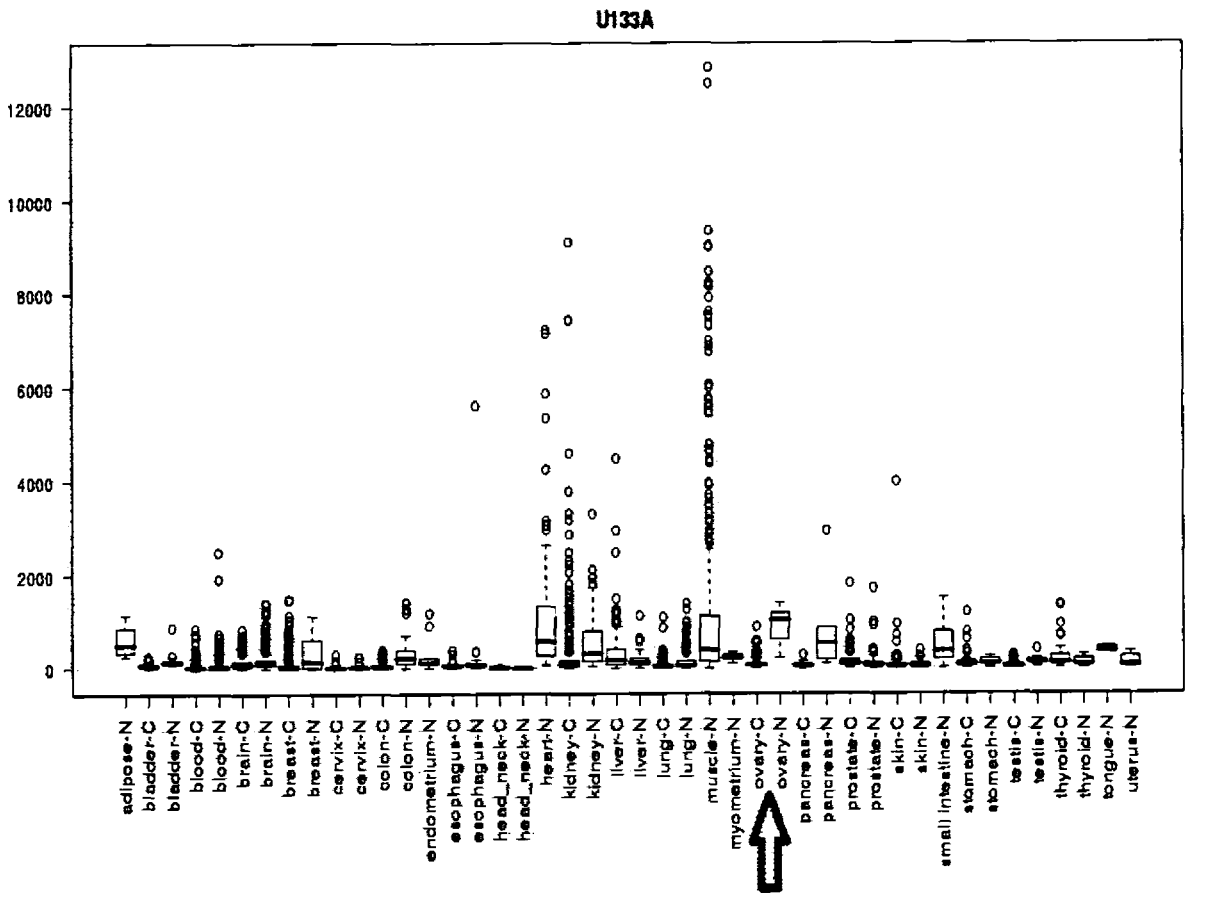
(a)


U133Plus2
35000
30000
25000
20000
15000
10000
5000
0
GIST-C
abdomin-C
adipose-C
adipose-N
adrenal gland-C
adrenal gland-N
bladder-C
bladder-N
blood-C
blood-N
brain-C
brain-N
breast-C
breast-N
cervix-C
cervix-N
colon-C
colon-N
endometrium-C
endometrium-N
esophagus-C
esophagus-N
head_neck-C
head_neck-N
kidney-C
kidney-N
liver-C
liver-N
liver-T
lung-C
lung-N
muscle-N
ovary-C
ovary-N
pancreas-C
pancreas-N
prostate-C
prostate-N
sarcoma
sarcoma-C
skin-C
skin-N
small intestine-C
small intestine-N
stomach-C
stomach-N
testis-C
testis-N
thyroid-C
thyroid-N
tongue-N
uterus-C
uterus-N
vagina-C
vagina-N
vulva-C
vulva-N

(b)


U133Plus2
log2
14
12
10
8
6
4
2
GIST-C
abdomin-C
adipose-C
adipose-N
adrenal gland-C
adrenal gland-N
bladder-C
bladder-N
blood-C
blood-N
brain-C
brain-N
breast-C
breast-N
cervix-C
cervix-N
colon-C
colon-N
endometrium-C
endometrium-N
esophagus-C
esophagus-N
head_neck-C
head_neck-N
kidney-C
kidney-N
liver-C
liver-N
liver-T
lung-C
lung-N
muscle-N
ovary-C
ovary-N
pancreas-C
pancreas-N
prostate-C
prostate-N
sarcoma
sarcoma-C
skin-C
skin-N
small intestine-C
small intestine-N
stomach-C
stomach-N
testis-C
testis-N
thyroid-C
thyroid-N
tongue-N
uterus-C
uterus-N
vagina-C
vagina-N
vulva-C
vulva-N

(c)


U133A
12000
10000
8000
6000
4000
2000
0
adipose-N
bladder-C
bladder-N
blood-C
blood-N
brain-C
brain-N
breast-C
breast-N
cervix-C
cervix-N
colon-C
colon-N
endometrium-N
esophagus-C
esophagus-N
head_neck-C
head_neck-N
heart-N
kidney-C
kidney-N
liver-C
liver-N
lung-C
lung-N
muscle-N
myometrium-N
ovary-C
ovary-N
pancreas-C
pancreas-N
prostate-C
prostate-N
skin-C
skin-N
small intestine-N
stomach-C
stomach-N
testis-C
testis-N
thyroid-C
thyroid-N
tongue-N
uterus-N

(d)

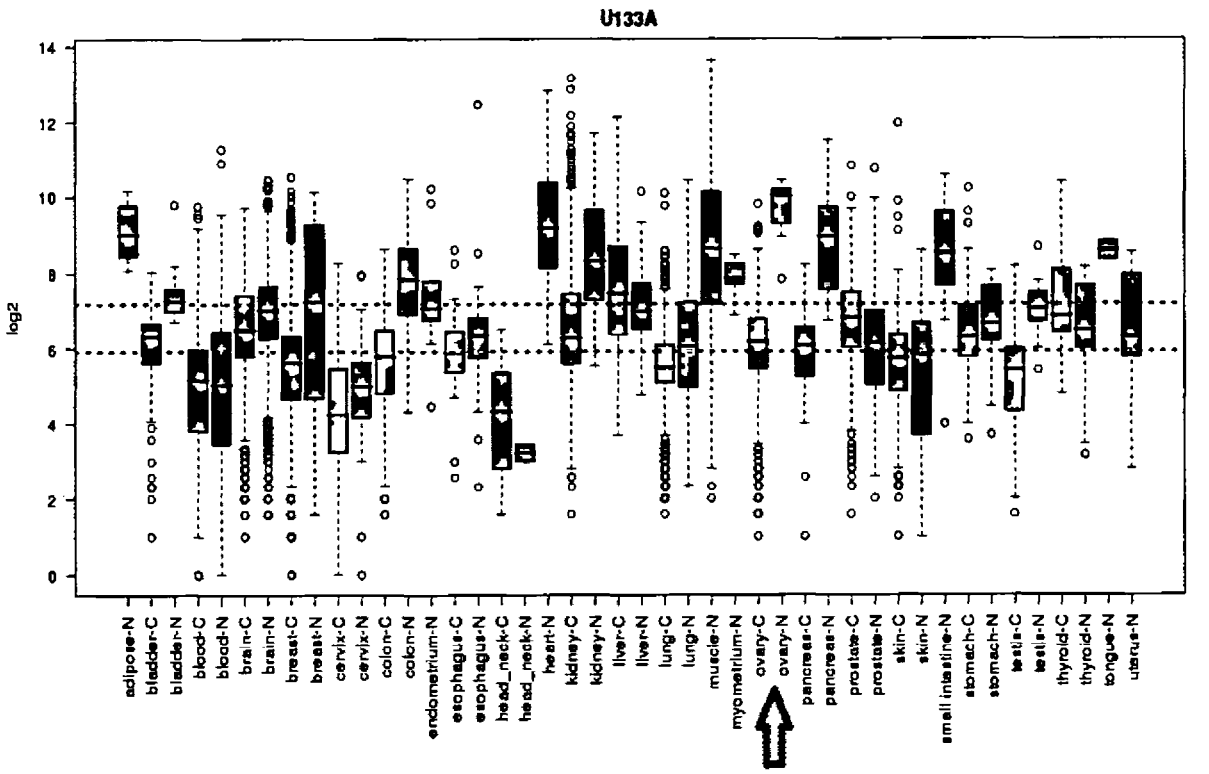
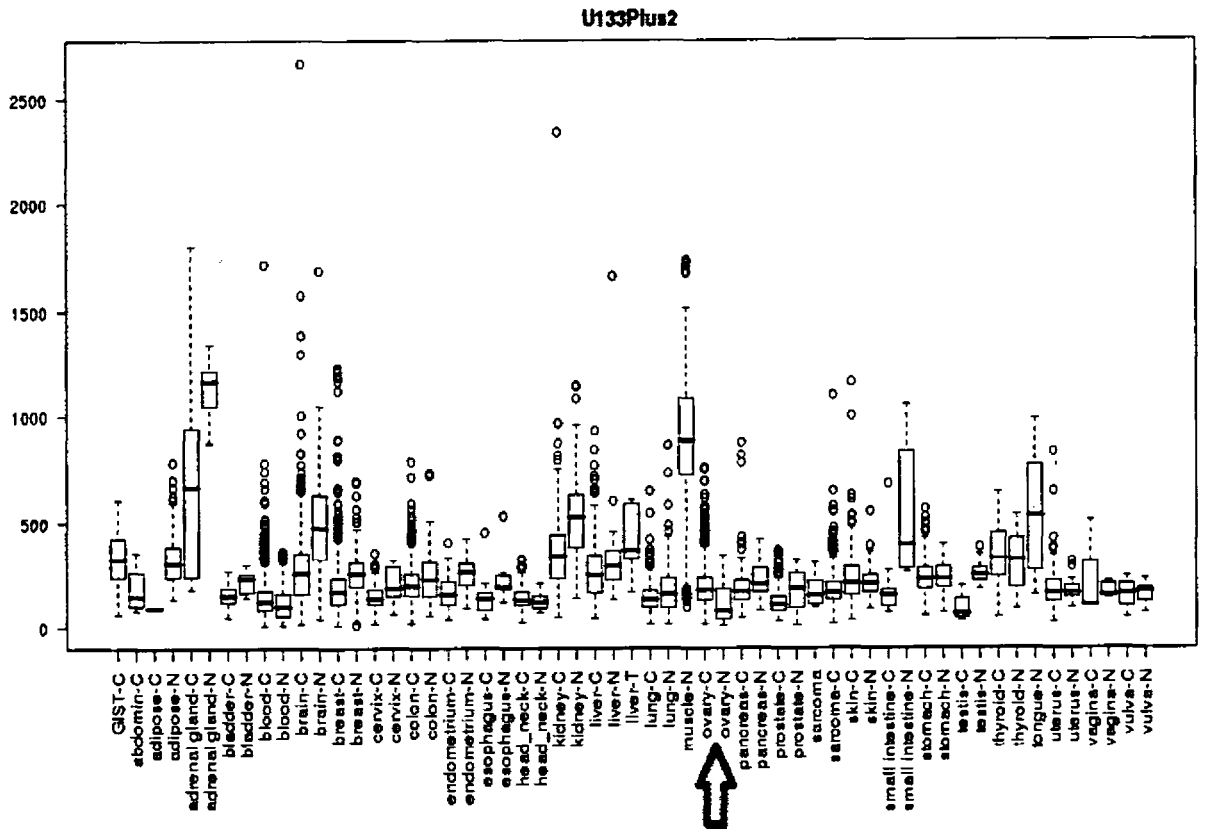


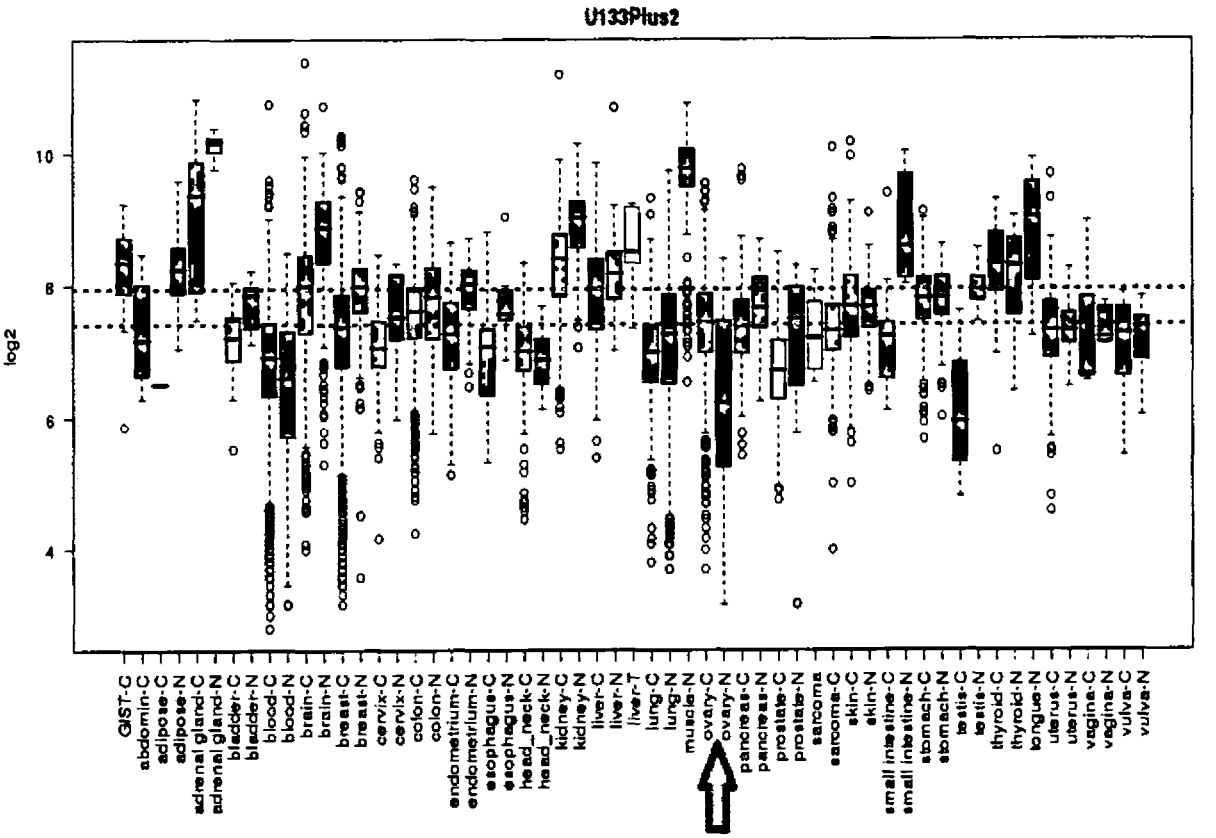
Figure 3.11: (a) and (b) Showing up regulation of PDK4 in data set U133Plus2, (c) and (d) Showing down regulation of PDK4 in dataset U133A

(Ovary-C = cancerous ovary, Ovary-N = normal ovary)

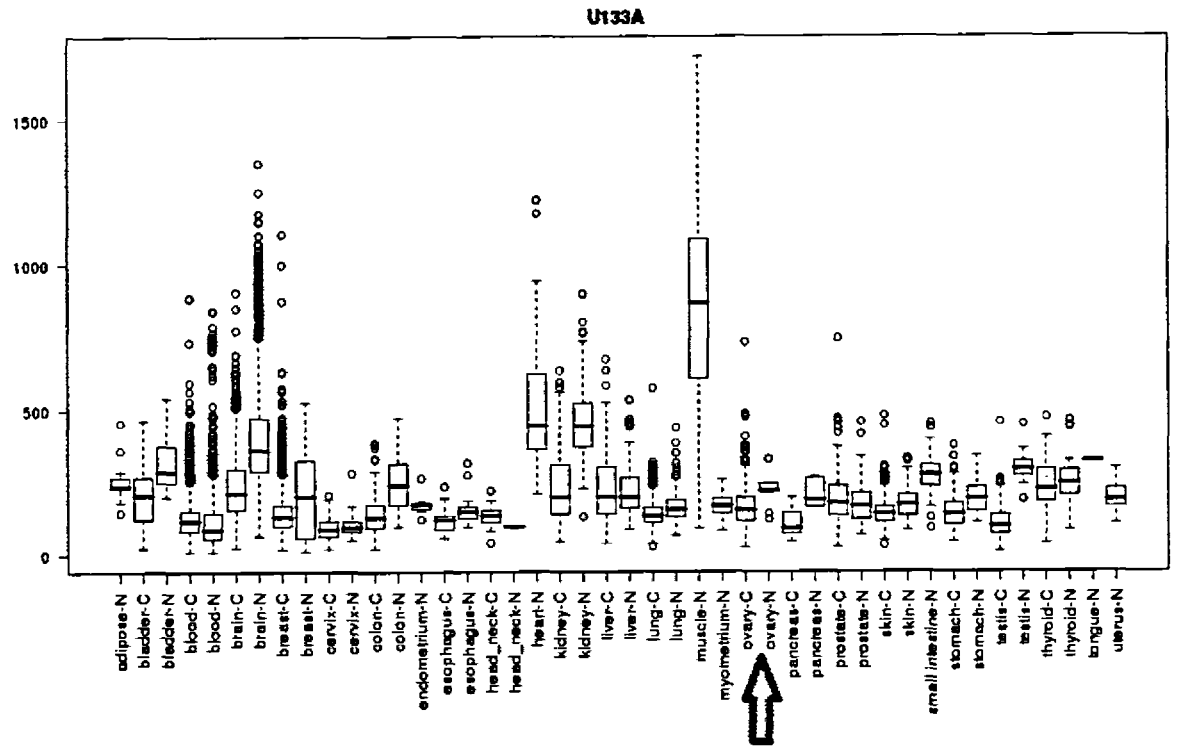
(a)


U133Plus2
2500
2000
1500
1000
500
0
GIST-C
abdomin-C
adipose-C
adipose-N
adrenal gland-C
adrenal gland-N
bladder-C
bladder-N
blood-C
blood-N
brain-C
brain-N
breast-C
breast-N
cervix-C
cervix-N
colon-C
colon-N
endometrium-C
endometrium-N
esophagus-C
esophagus-N
head_neck-C
head_neck-N
kidney-C
kidney-N
liver-C
liver-N
liver-T
lung-C
lung-N
muscle-N
ovary-C
ovary-N
pancreas-C
pancreas-N
prostate-C
prostate-N
sarcoma
sarcoma-C
skin-C
skin-N
small intestine-C
small intestine-N
stomach-C
stomach-N
testis-C
testis-N
thyroid-C
thyroid-N
tongue-N
uterus-C
uterus-N
vagina-C
vagina-N
vulva-C
vulva-N

(b)


U133Plus2
log2
10
8
6
4
GIST-C
abdomin-C
adipose-C
adipose-N
adrenal gland-C
adrenal gland-N
bladder-C
bladder-N
blood-C
blood-N
brain-C
brain-N
breast-C
breast-N
cervix-C
cervix-N
colon-C
colon-N
endometrium-C
endometrium-N
esophagus-C
esophagus-N
head_neck-C
head_neck-N
kidney-C
kidney-N
liver-C
liver-N
liver-T
lung-C
lung-N
muscle-N
ovary-C
ovary-N
pancreas-C
pancreas-N
prostate-C
prostate-N
sarcoma
sarcoma-C
skin-C
skin-N
small intestine-C
small intestine-N
stomach-C
stomach-N
testis-C
testis-N
thyroid-C
thyroid-N
tongue-N
uterus-C
uterus-N
vagina-C
vagina-N
vulva-C
vulva-N

(c)


U133A
1500
1000
500
0
adipose-N
bladder-C
bladder-N
blood-C
blood-N
brain-C
brain-N
breast-C
breast-N
cervix-C
cervix-N
colon-C
colon-N
endometrium-N
esophagus-C
esophagus-N
head_neck-C
head_neck-N
heart-N
kidney-C
kidney-N
liver-C
liver-N
lung-C
lung-N
muscle-N
myometrium-N
ovary-C
ovary-N
pancreas-C
pancreas-N
prostate-C
prostate-N
skin-C
skin-N
small intestine-N
stomach-C
stomach-N
testis-C
testis-N
thyroid-C
thyroid-N
tongue-N
uterus-N

(d)

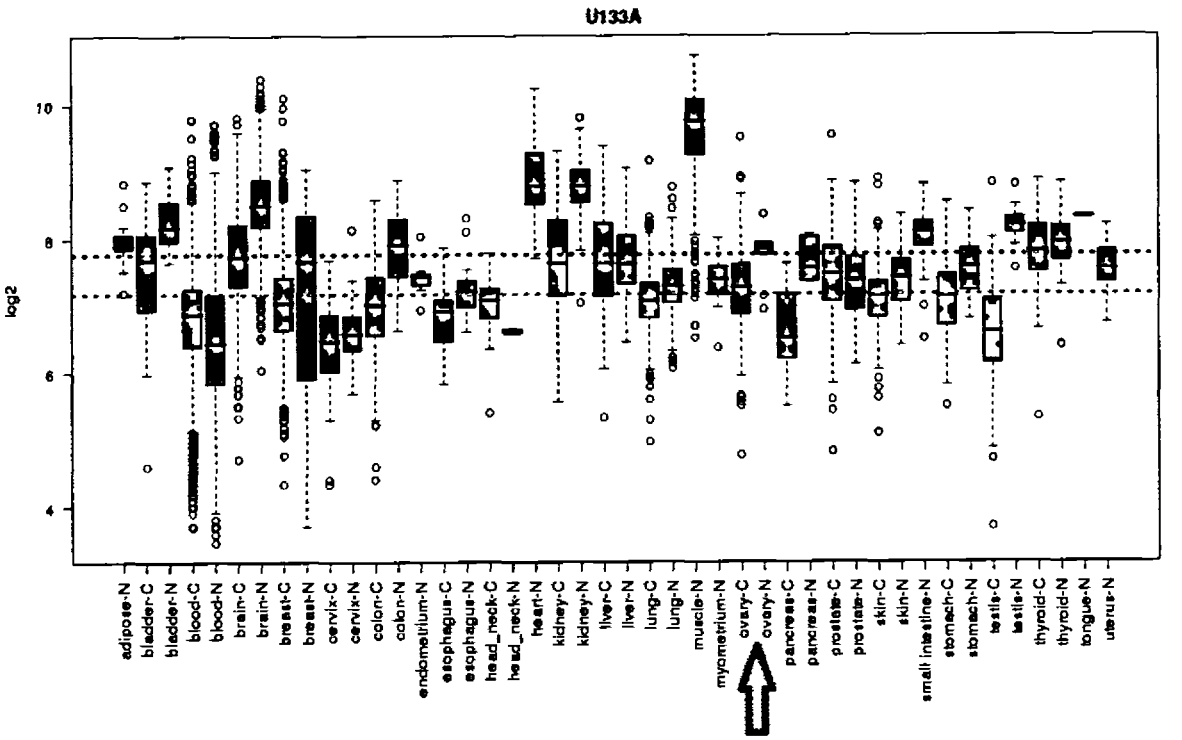


Figure 3.12: Showing up regulation of PDK2 in (a) and (b) Data set U133A (c) and (d) Data set U133A

(Ovary-C = cancerous ovary, Ovary-N = normal ovary)

Table 3.2 Showing p-values and fold change of PDK2 and PDK4 in both datasets

Reference ID	Gene Symbol	GSE4122		GSE6008	
		P value	Fold change	P value	Fold change
202590_s_at	PDK2	2.47E-05	1.08E+31	0.006889	1.1428 (non significant)
205960_at	PDK4	0.005181	4.16E+15	4.58E-13	2.064555

Table 3.3 SAGE results for PDK2 and PDK4

Gene Symbol	Unigene Cluster	SAGE Tag	Normal (*tpm)		Ovarian Cancer (*tpm)	
			Tissues	Cell lines	Tissues	Cell lines
PDK2	Hs.256667	TGTGCTCAGG	No data	0	6	6
PDK4	Hs.8364	TAAATACTTG	No data	0	0	4

*Tags per million

Table 3.4 Common microRNA's predicted by TargetScan, Pictar, miRanda and miRGen along with their binding energies.

	MicroRNA's	Chromosomal Location	Binding Energy
PDK4	hsa-miR-497	17p13.1	-116.9 kcal/mol
	hsa-miR-424	Xq26.3	-79.7 kcal/mol
	hsa-miR-15b	3q25.33	-58.5 kcal/mol
	hsa-miR-15a	13q14.2	-62.9 kcal/mol
PDK2	hsa-miR-195	17p13.1	-85.6 kcal/mol
	hsa-miR-16	13q14.2	-64.5 kcal/mol
	hsa-miR-326	11q13.4	-94.9 kcal/mol
	hsa-miR-330-5p	19q13.32	-27.7 kcal/mol
	hsa-miR-1208	8q24.21	-74.3 kcal/mol

3.2 Claudin Family

The microarray analysis of gene expression data sets from ovarian samples evidently provided the proof of involvement of claudin gene family in human ovarian cancer. The differential expression of cldn3, cldn4 and cldn7 shows the full participation of this family in cancer progression. So in order to completely understand the behavior of claudin family in cancer development the whole claudin family was analyzed separately.

For this expression data of only cluadin genes present in GSE6008 and GSE4122 was separated. The small datasets only for claudin gene family was made. These data sets were used for further analysis. The results obtained from both datasets (Table 3.1) showed that cldn4, cldn7, cldn16 and cldn3 are highly over-expressed in ovarian cancer while cldn5 is down regulated. Cldn6 and cldn15 showed a very different behavior, as cldn6 is found to have over expression and down regulation of cldn15 in dataset GSE4122 while in GSE6008 no significant difference was detected. These findings through fold change analysis were further verified through t-test and SAM.

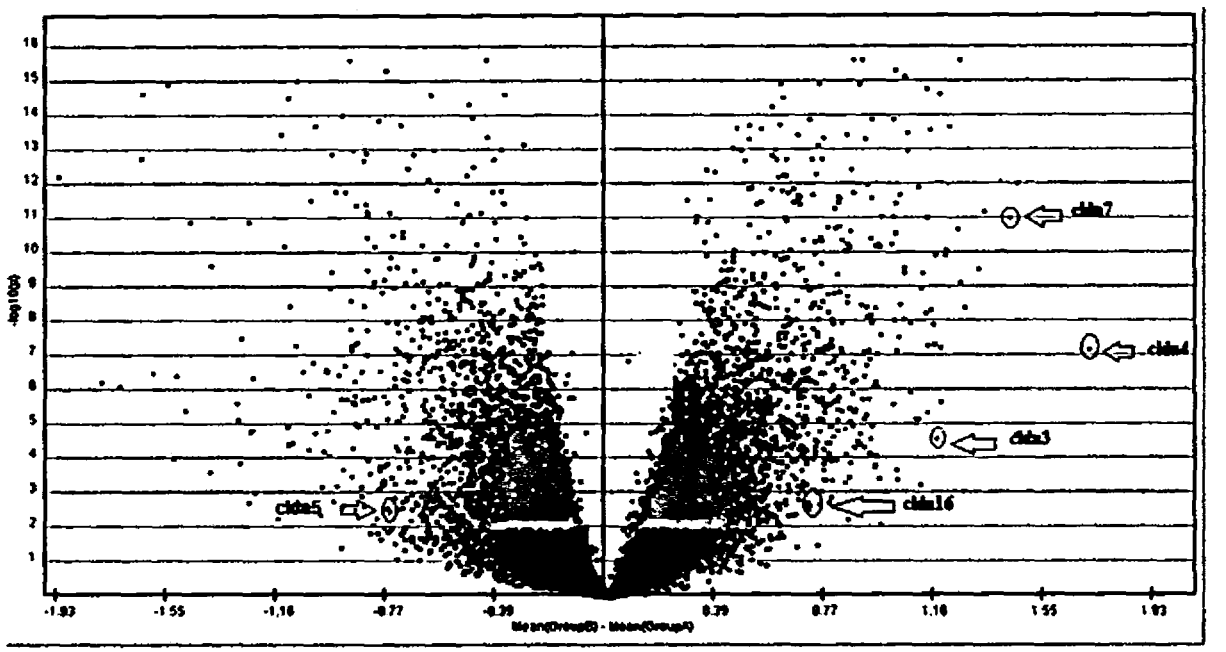
The t-test confirmed the cldn3, cldn4, cldn7, cldn16, cldn15 and cldn5 as significant genes and cldn6 was the only non significant gene so it was excluded from further analysis (Table 3.5). The differential expression of these significant genes was also detected in a volcano plot mentioned in Figure 3.13. Further mining of selected members of claudin family was done through SAM which separated cldn4, cldn3, and cldn7 as positive significant genes and cldn5 as negative significant, shown in SAM graph Figure 3.14. Same results were also examined by hierarchical clustering, k-means clustering and self organizing tree algorithm (Figure 3.15 and Figure 3.16). SAGE analysis was done in order to confirm the observed

results (Table 3.6). For the visual description of SAGE results Sage map was used which also confirmed the over expression of cldn3, cldn4 and cldn7 (Figure 3.17, 3.18 and 3.19 respectively) but very negligible up regulation of cldn16 (Figure 3.20) and surprisingly cldn5 showed up regulation as compared to microarray results (Figure 3.21).

The results obtained from SAGE were contradictory to microarray results so we also checked the behavior of genes was checked using GENT. The over expression of cldn3, cldn4 and cldn7 was also confirmed by GENT (Figure 3.22, 3.23 and 3.24 respectively). Here GENT confirms the up regulation of cldn16 and it also shows that in cldn16 was highly expressed in ovarian cancer than any other cancer (Figure 3.25). In case of cldn5 GENT agreed with the SAGE by showing clearly the over expression in ovarian cancer samples (Figure 3.26).

The microRNA targets were then predicted for only those claudin members which were proved to be associated with ovarian cancer. Common results from all of microRNA prediction tools were noted as the most reliable targets (Table 3.7). As these micorRNA's were most susceptible candidates for being involved in ovarian cancer by regulating the expression of claudin genes.

(a)



(b)

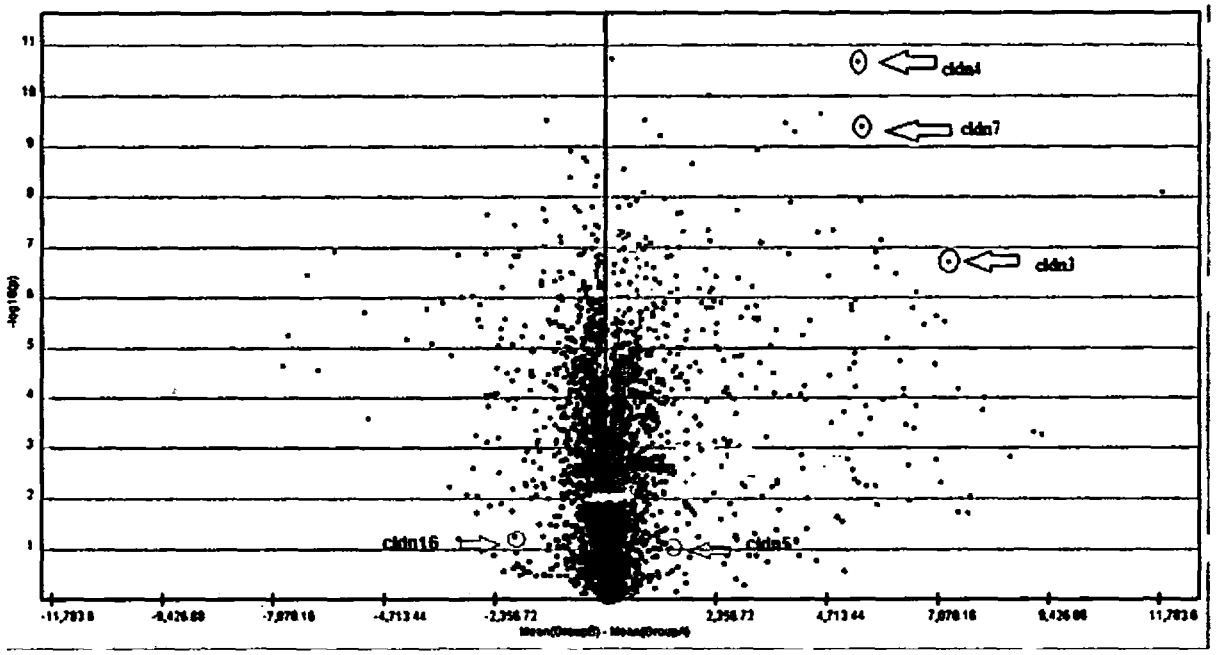
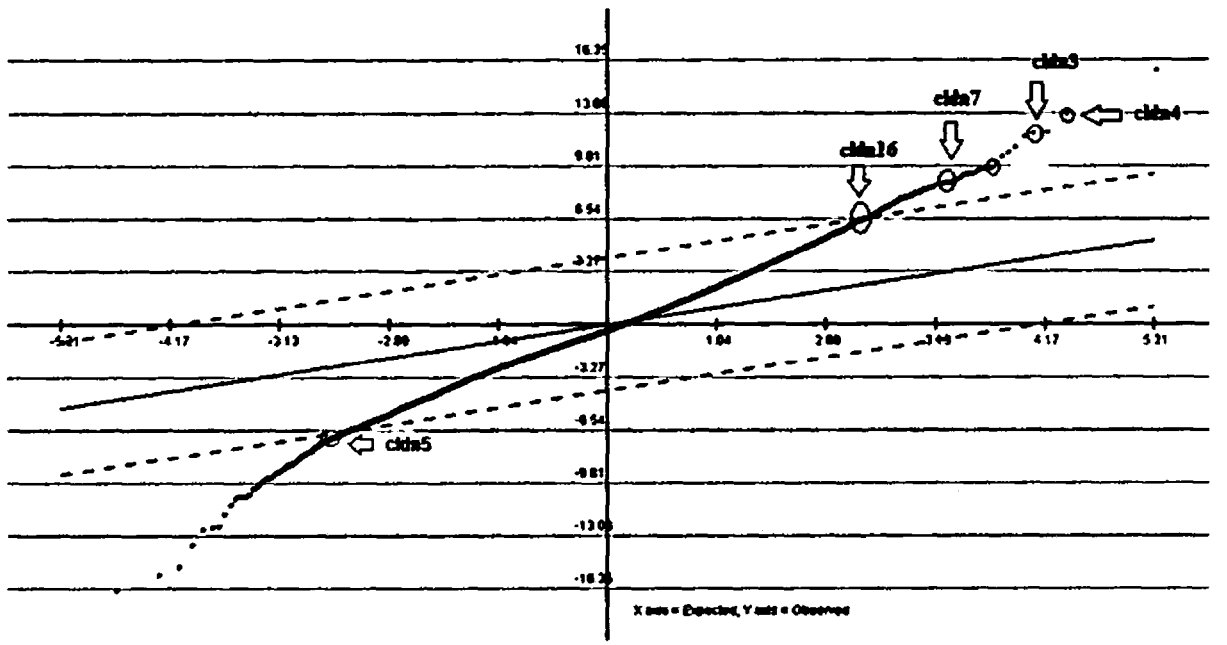


Figure 3.13: Volcano plot for claudin genes (a) GSE6008 (b) GSE4122

(a)



(b)

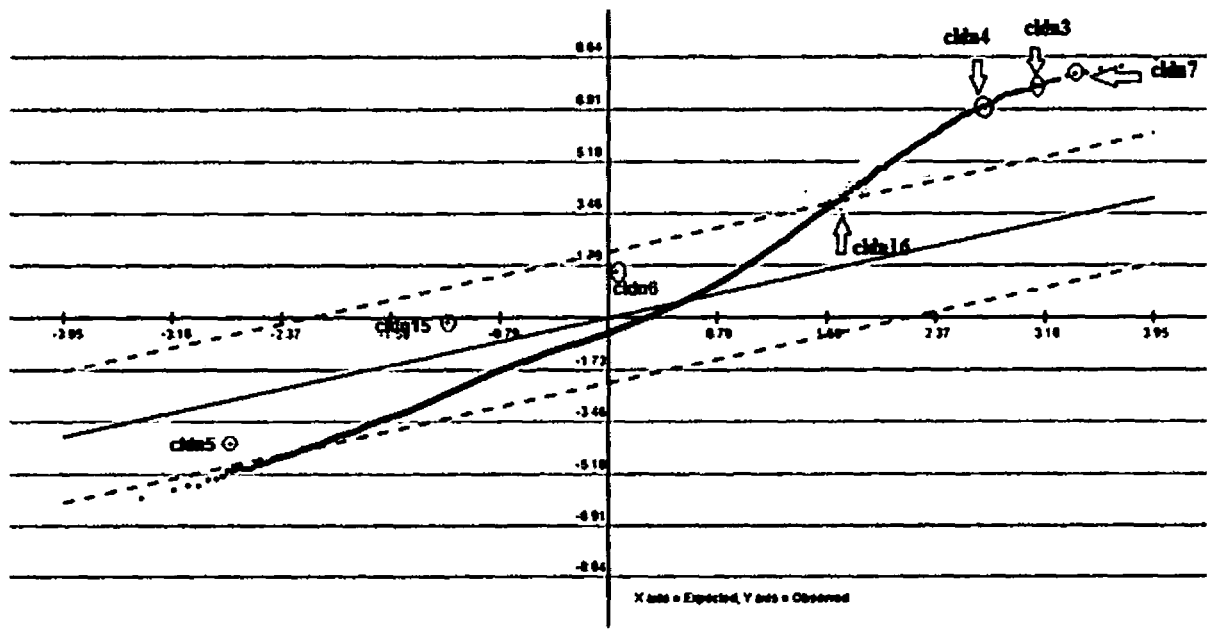
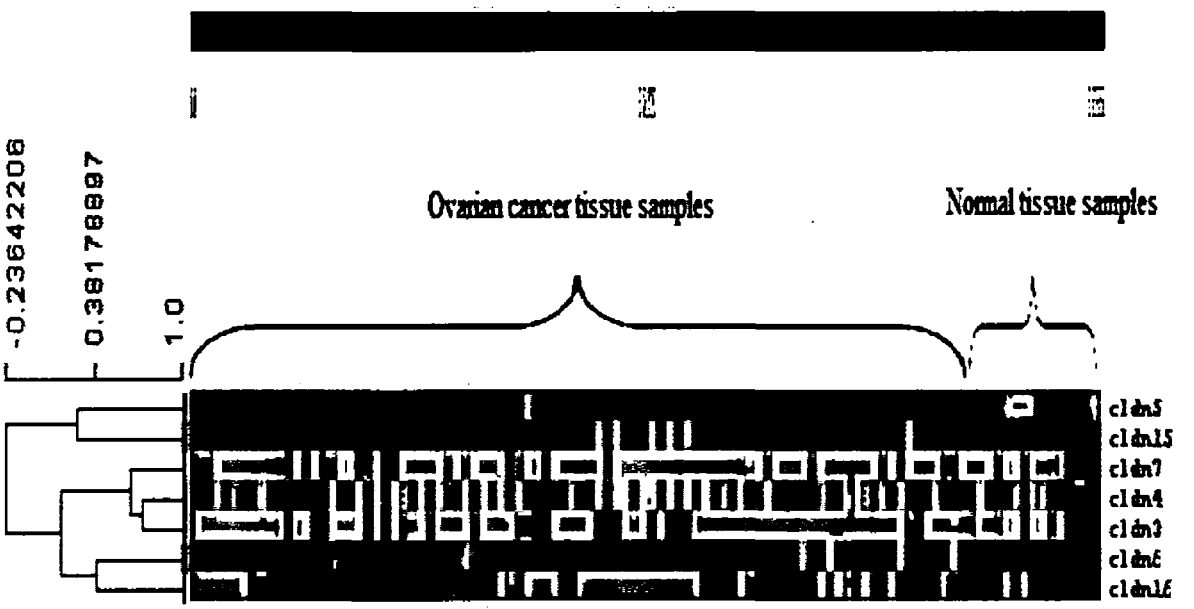
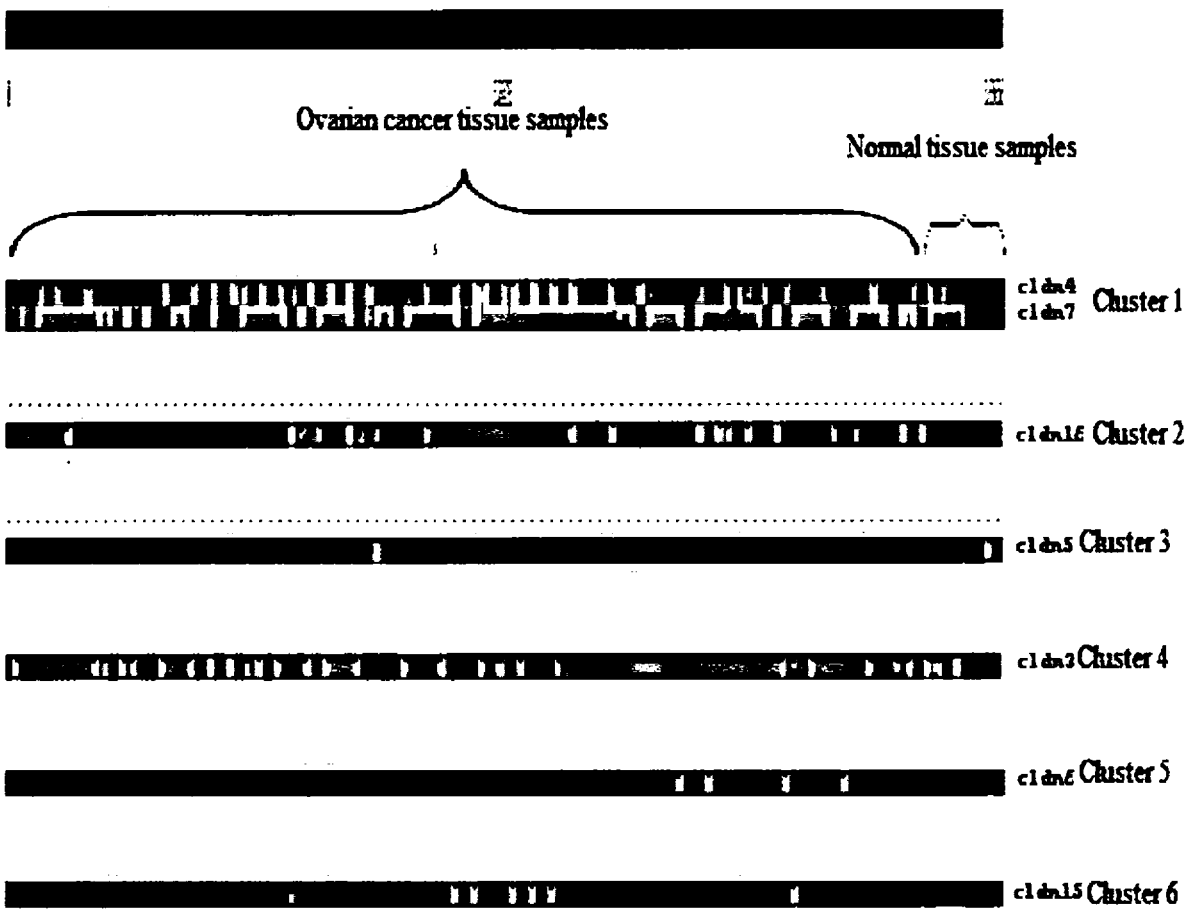


Figure 3.14: SAM graph for claudins (a) GSE6008 (b) GSE4122

(a)


-0.23642206
0.38178887
1.0
Ovarian cancer tissue samples
Normal tissue samples

(b)


Ovarian cancer tissue samples
Normal tissue samples
c1dn4
c1dn7
Cluster 1
c1dn16
Cluster 2
c1dn5
Cluster 3
c1dn3
Cluster 4
c1dn6
Cluster 5
c1dn15
Cluster 6

(c)

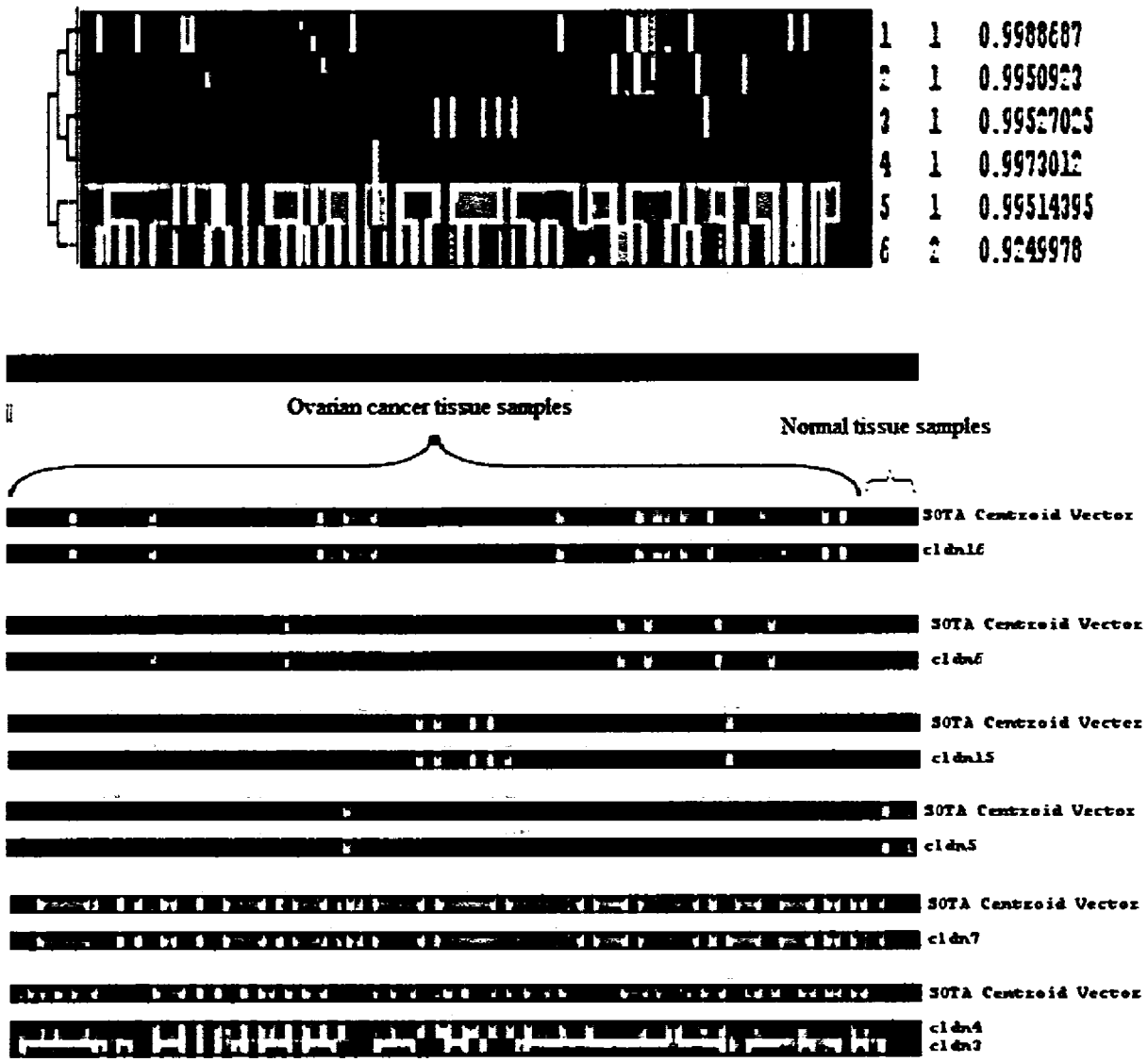
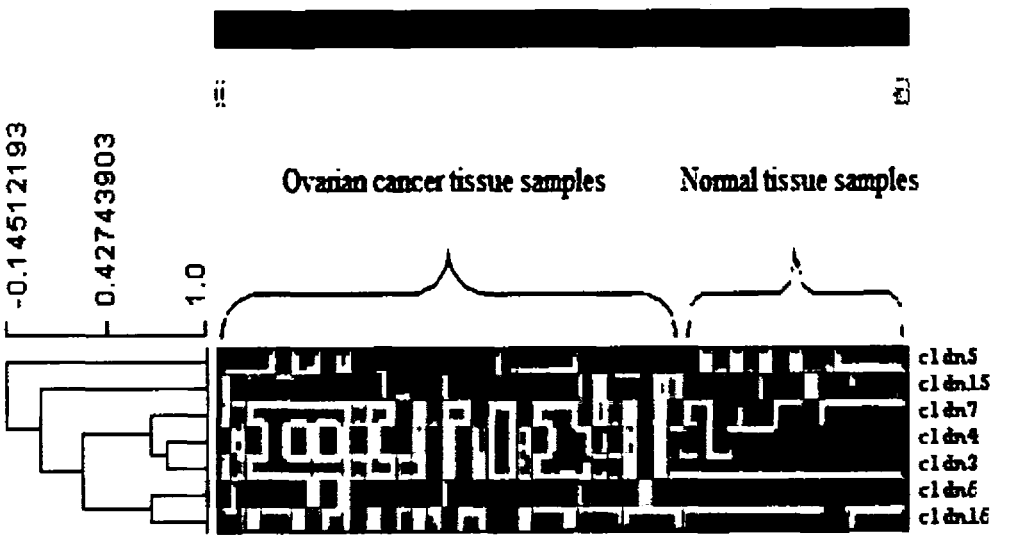
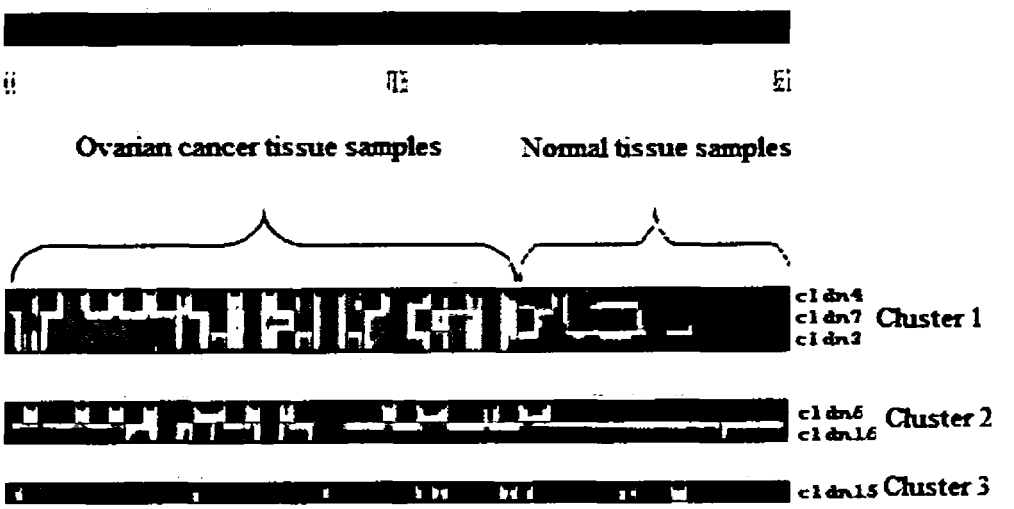


Figure 3.15: Clustering diagrams of PDK2 from data set GSE6008 (a) Hierarchical cluster (b) K-means clusters (c) SOTA cluster

(a)


-0.14512193
0.42743903
1.0
Ovarian cancer tissue samples
Normal tissue samples
cl dn5
cl dn15
cl dn7
cl dn4
cl dn2
cl dn6
cl dn16

(b)


Ovarian cancer tissue samples
Normal tissue samples
cl dn4
cl dn7
cl dn2
Cluster 1
cl dn6
cl dn16
Cluster 2
cl dn15
Cluster 3

(c)

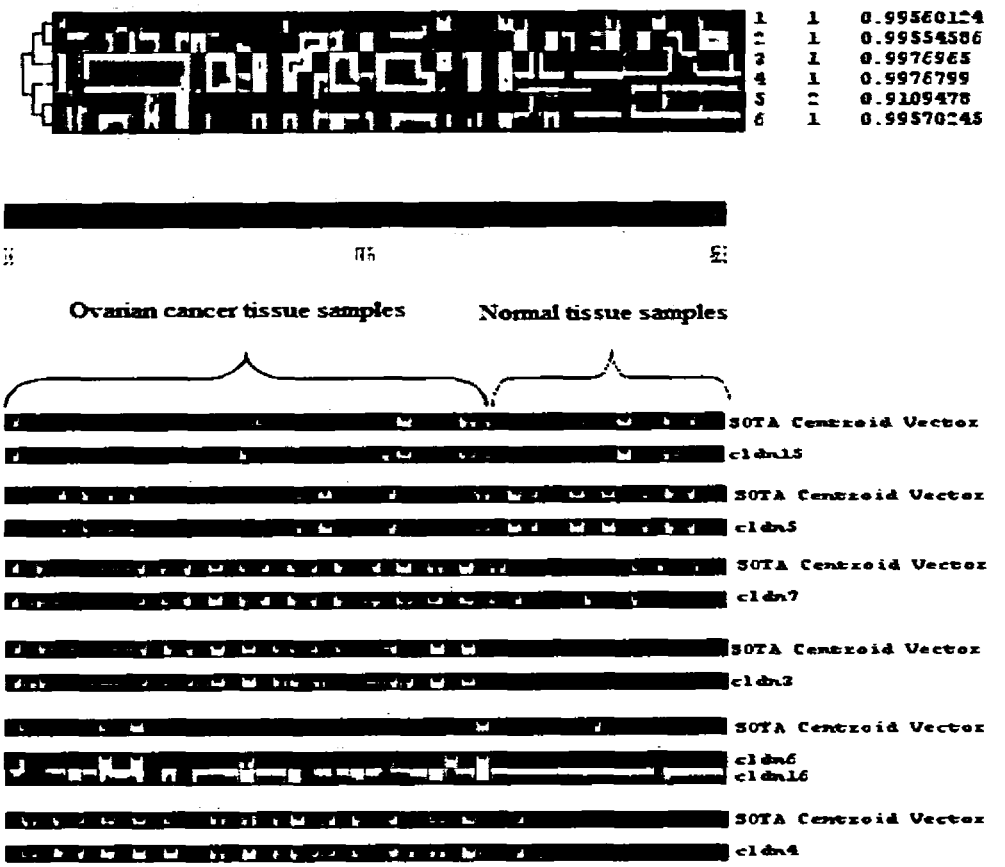


Figure 3.16: Clustering diagrams of claudins from data set GSE4122 (a) Hierarchical cluster (b) K-means clusters (c) SOTA cluster

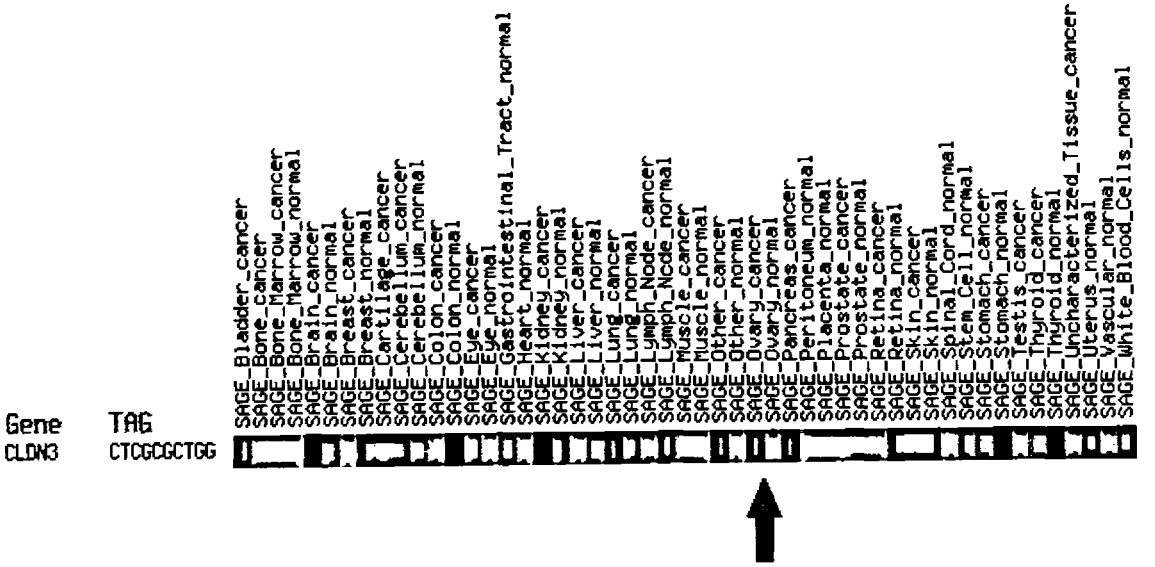


Figure 3.17: Short SAGE map of cldn3 from Gene Finder tool showing over expression in cancerous ovary.

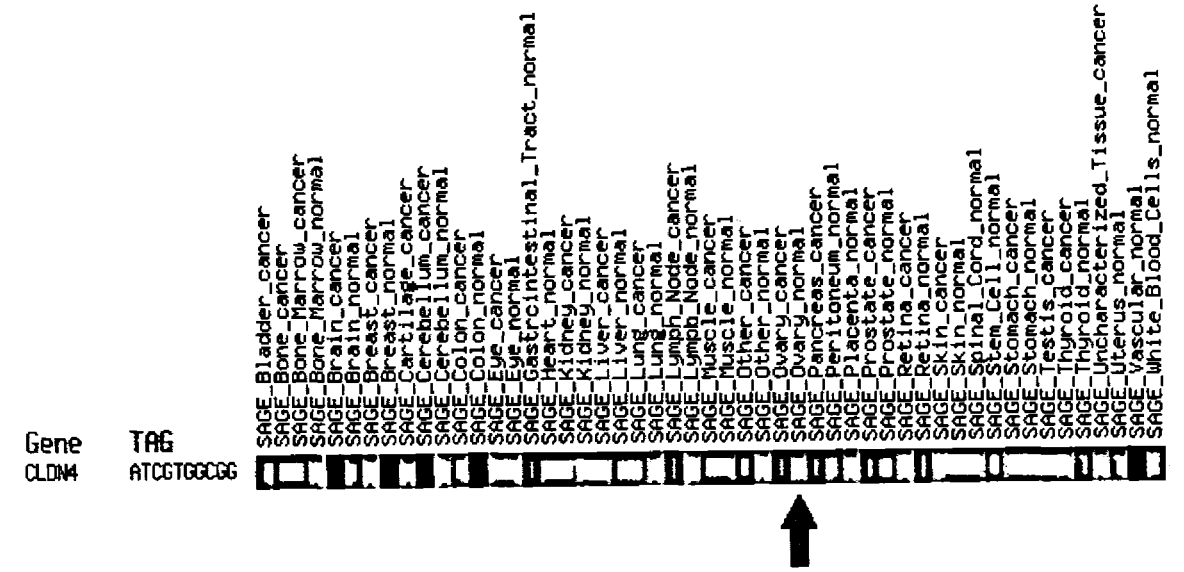


Figure 3.18: Short SAGE map of cldn4 from Gene Finder tool showing over expression in cancerous ovary.

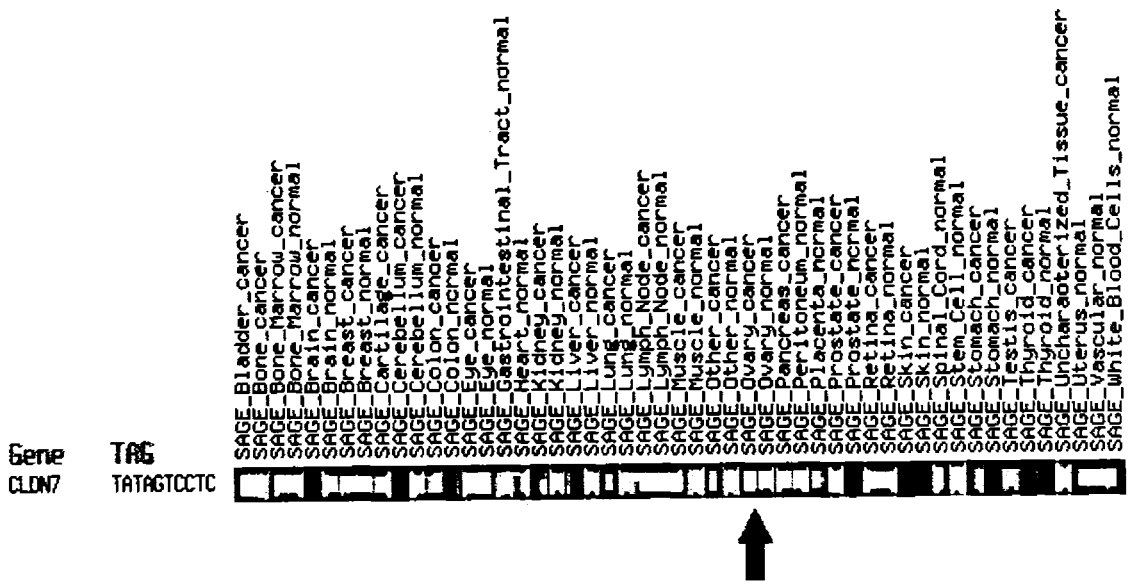


Figure 3.19: Short SAGE map of cldn7 from Gene Finder tool showing over expression in cancerous ovary.

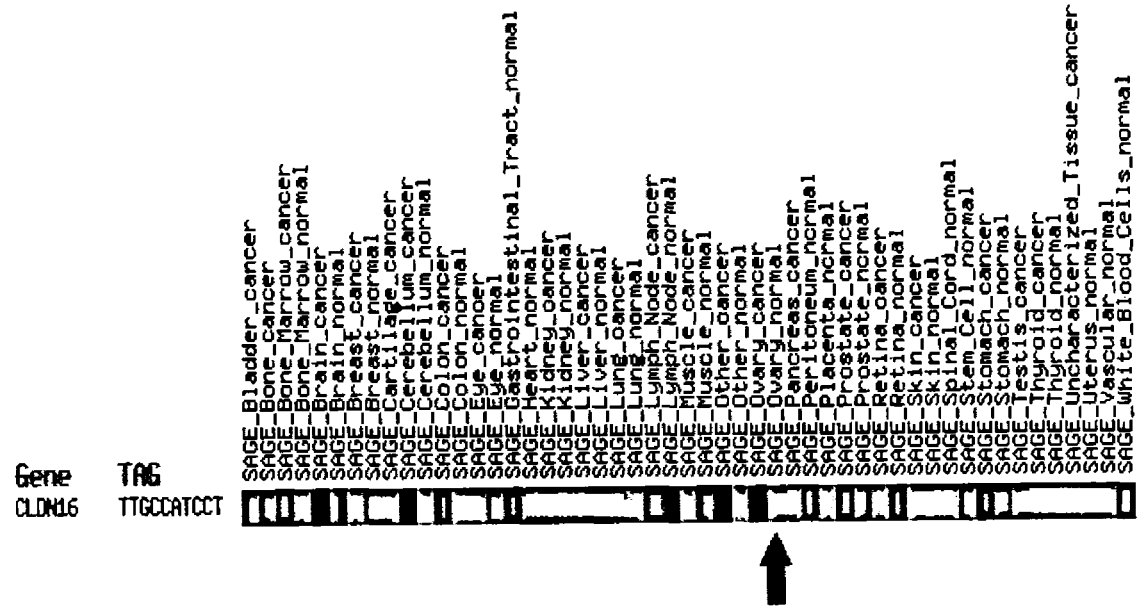


Figure 3.20: Short SAGE map of cldn16 from Gene Finder tool showing over expression in cancerous ovary.

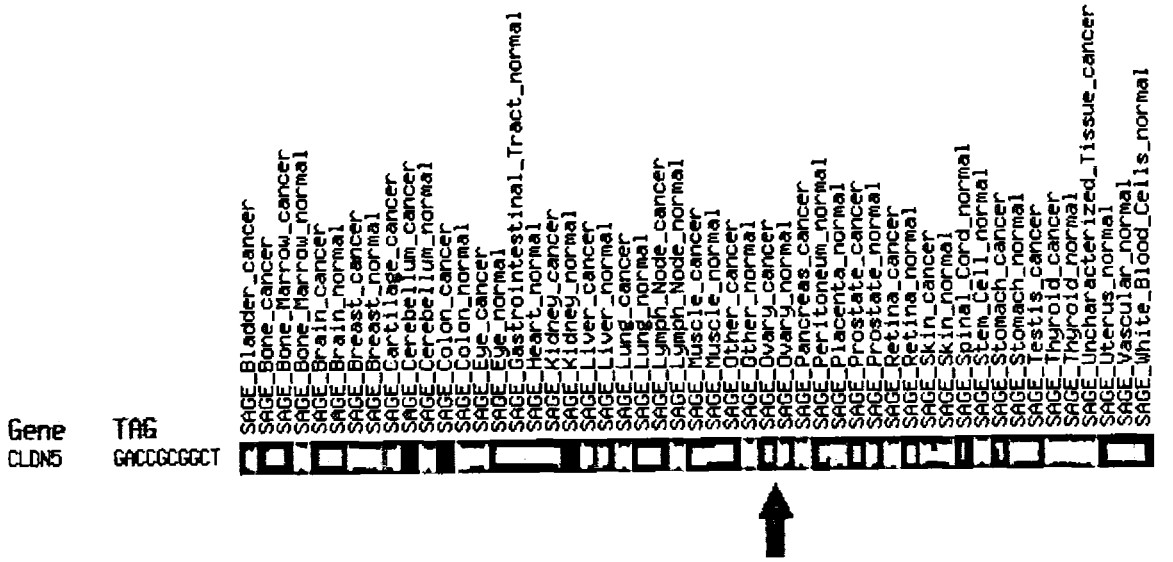
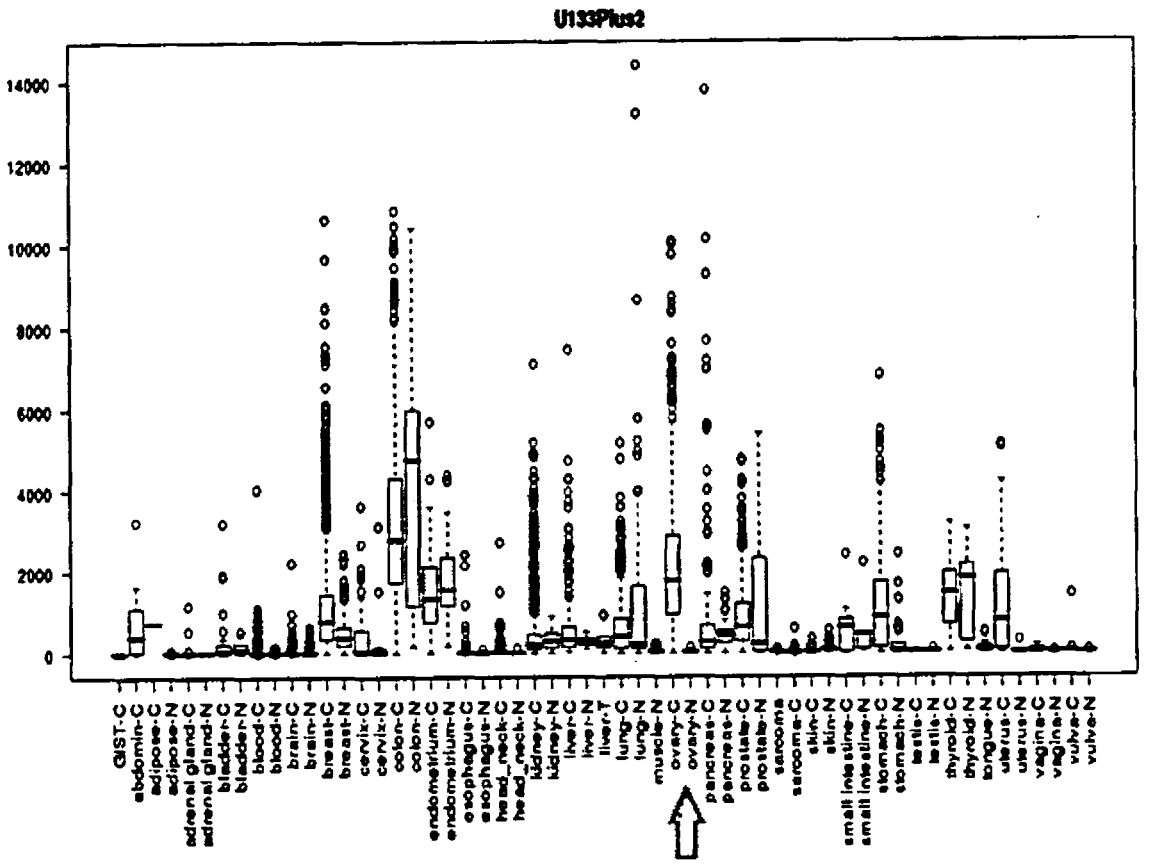
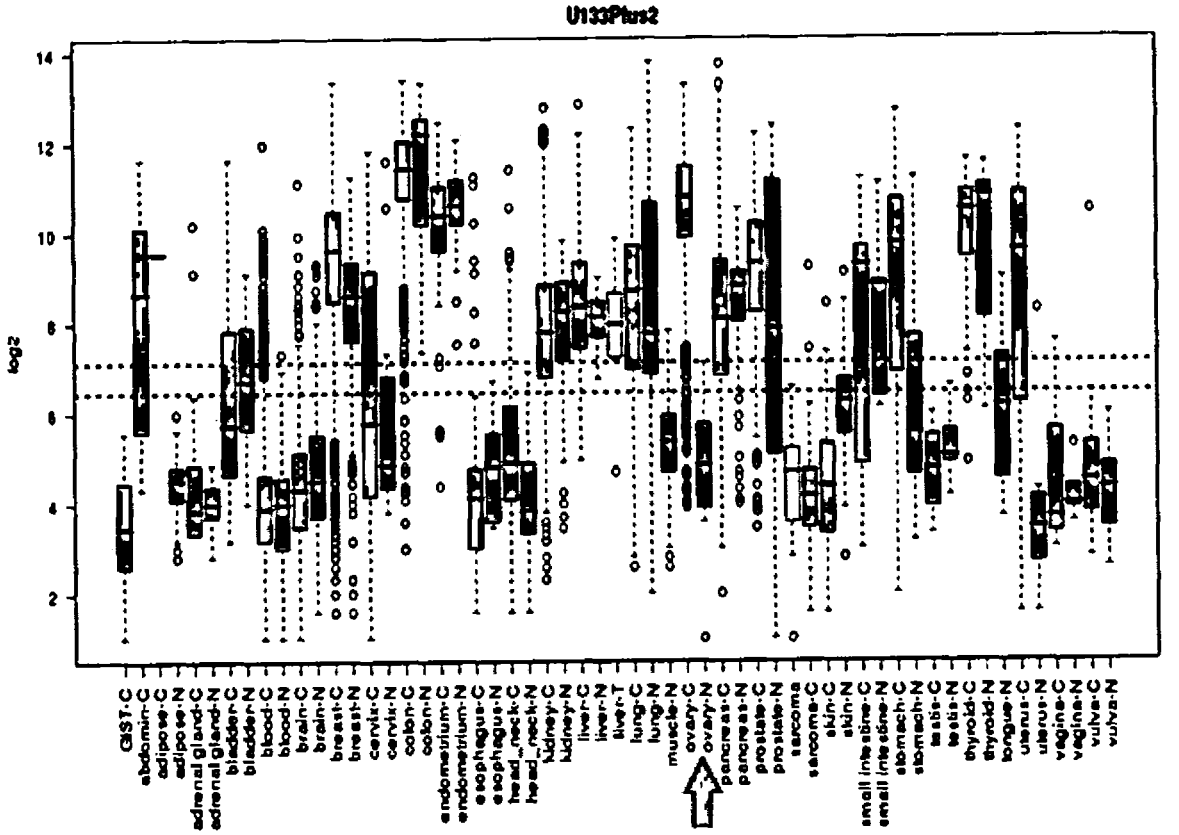


Figure 3.21: Short SAGE map of cldn5 from Gene Finder tool showing over expression in cancerous ovary.

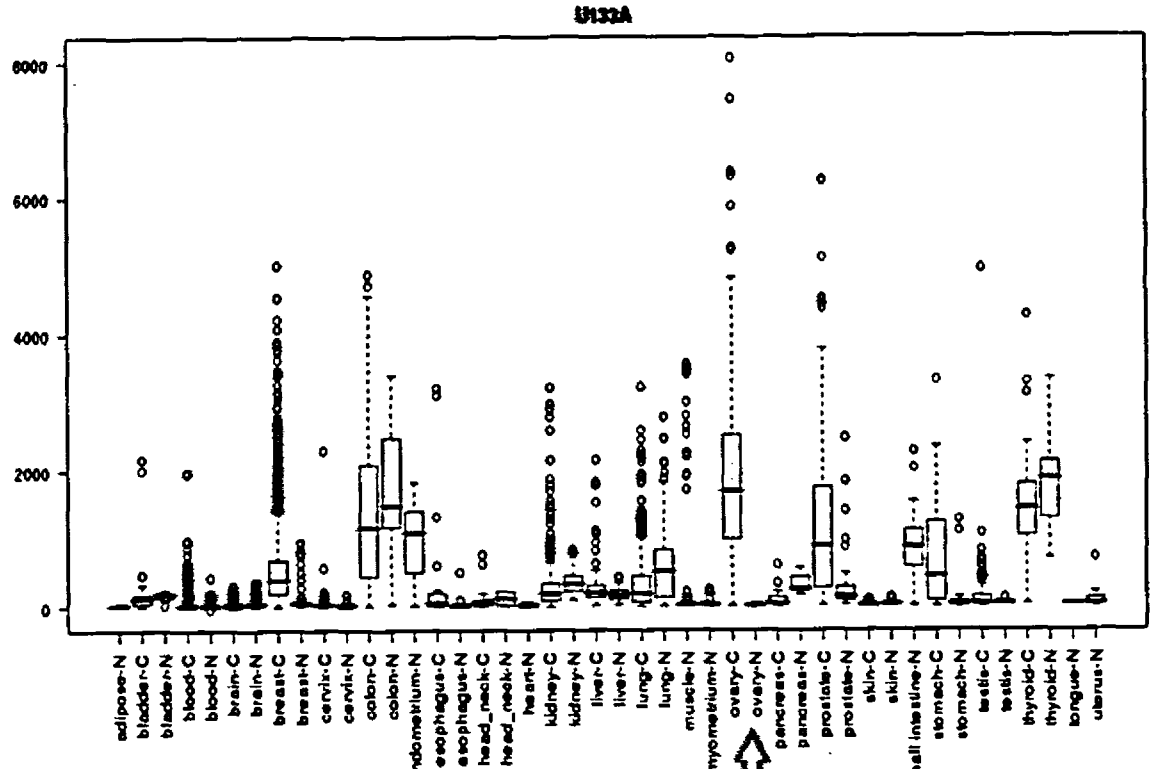
(a)


U133Plus2
14000
12000
10000
8000
6000
4000
2000
0
GIST-C
abdomin-C
adipose-C
adipose-N
adrenal gland-C
adrenal gland-N
bladder-C
bladder-N
blood-C
blood-N
brain-C
brain-N
breast-C
breast-N
cervix-C
cervix-N
colon-C
colon-N
endometrium-C
endometrium-N
esophagus-C
esophagus-N
head_neck-C
head_neck-N
kidney-C
kidney-N
liver-C
liver-N
lung-C
lung-N
muscle-N
ovary-C
ovary-N
pancreas-C
pancreas-N
prostate-C
prostate-N
sarcoma
sarcoma-C
skin-C
skin-N
small intestine-C
small intestine-N
stomach-C
stomach-N
testis-C
testis-N
thyroid-C
thyroid-N
tongue-N
uterus-C
uterus-N
vagina-C
vagina-N
vulva-C
vulva-N

(b)


U133Plus2
log2
14
12
10
8
6
4
2
GIST-C
abdomin-C
adipose-C
adipose-N
adrenal gland-C
adrenal gland-N
bladder-C
bladder-N
blood-C
blood-N
brain-C
brain-N
breast-C
breast-N
cervix-C
cervix-N
colon-C
colon-N
endometrium-C
endometrium-N
esophagus-C
esophagus-N
head_neck-C
head_neck-N
kidney-C
kidney-N
liver-C
liver-N
liver-T
lung-C
lung-N
muscle-N
ovary-C
ovary-N
pancreas-C
pancreas-N
prostate-C
prostate-N
sarcoma
sarcoma-C
skin-C
skin-N
small intestine-C
small intestine-N
stomach-C
stomach-N
testis-C
testis-N
thyroid-C
thyroid-N
tongue-N
uterus-C
uterus-N
vagina-C
vagina-N
vulva-C
vulva-N

(c)


U133A
8000
6000
4000
2000
0
adipose-N
bladder-C
bladder-N
blood-C
blood-N
brain-C
brain-N
breast-C
breast-N
cervix-C
cervix-N
colon-C
colon-N
endometrium-N
esophagus-C
esophagus-N
head_neck-C
head_neck-N
heart-N
kidney-C
kidney-N
liver-C
liver-N
lung-C
lung-N
muscle-N
myometrium-N
ovary-C
ovary-N
pancreas-C
pancreas-N
prostate-C
prostate-N
skin-C
skin-N
small intestine-N
stomach-C
stomach-N
testis-C
testis-N
thyroid-C
thyroid-N
tongue-N
uterus-N

(d)

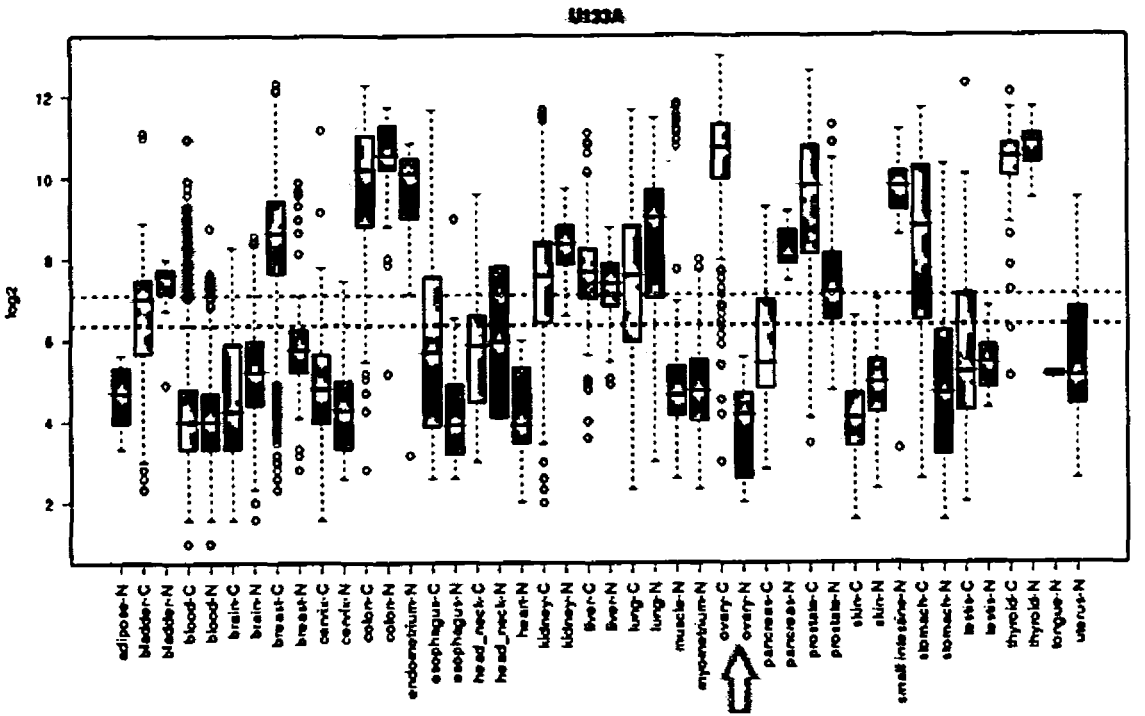
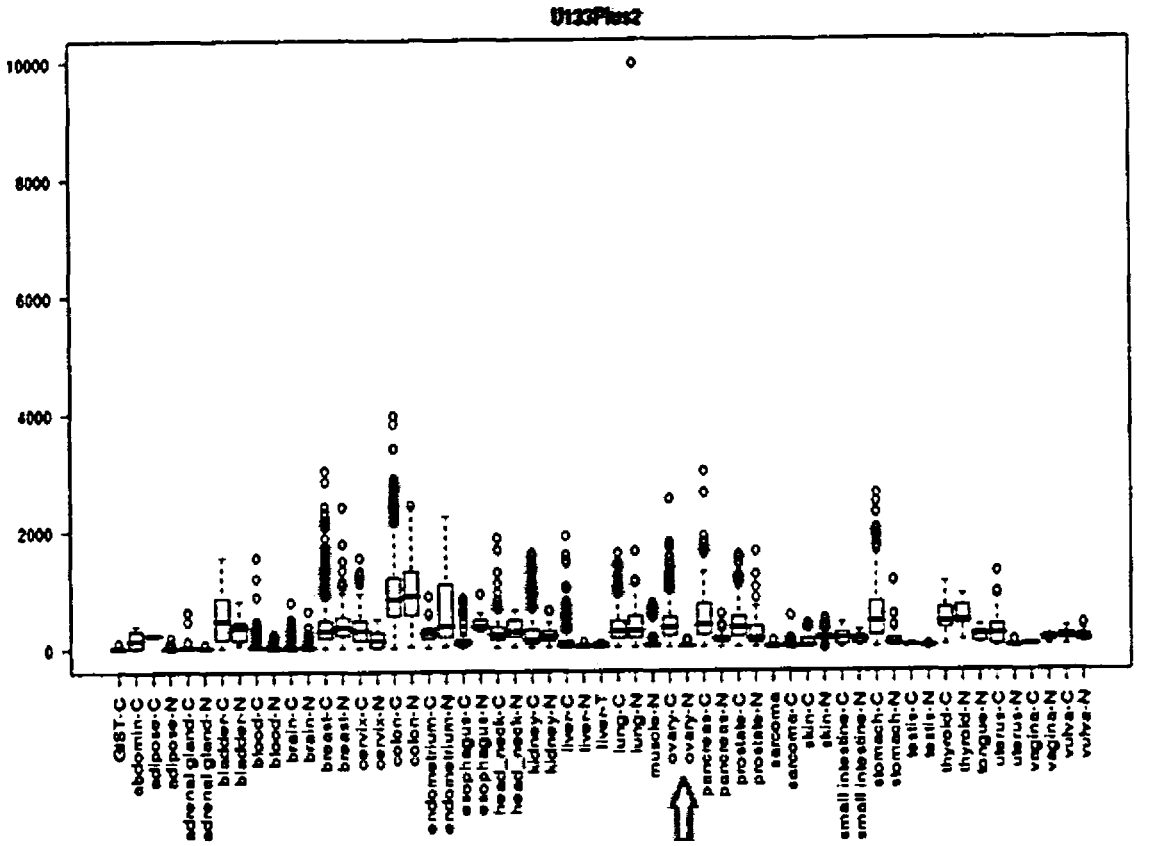


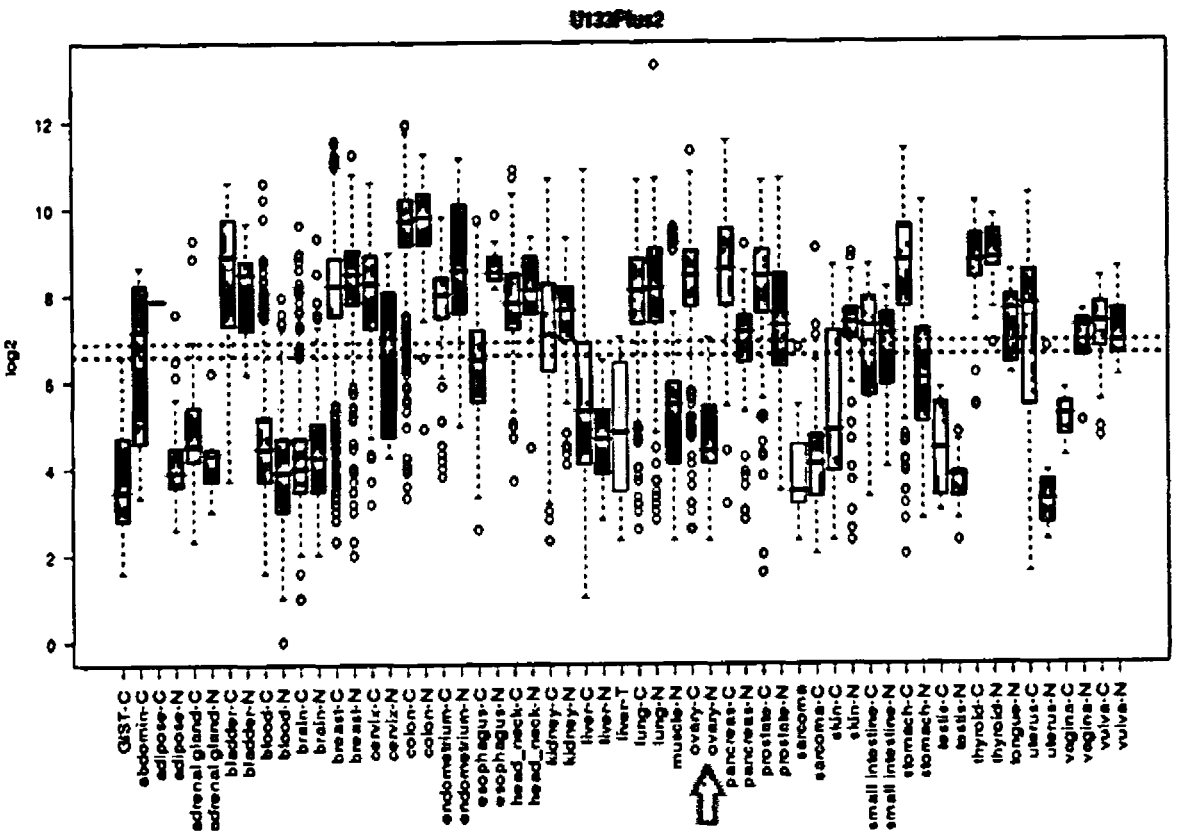
Figure 3.22: Showing up regulation of Cldn3 (a) and (b) Data set U133A (c) and (d) Data

(Ovary-C = cancerous ovary, Ovary-N = normal ovary)

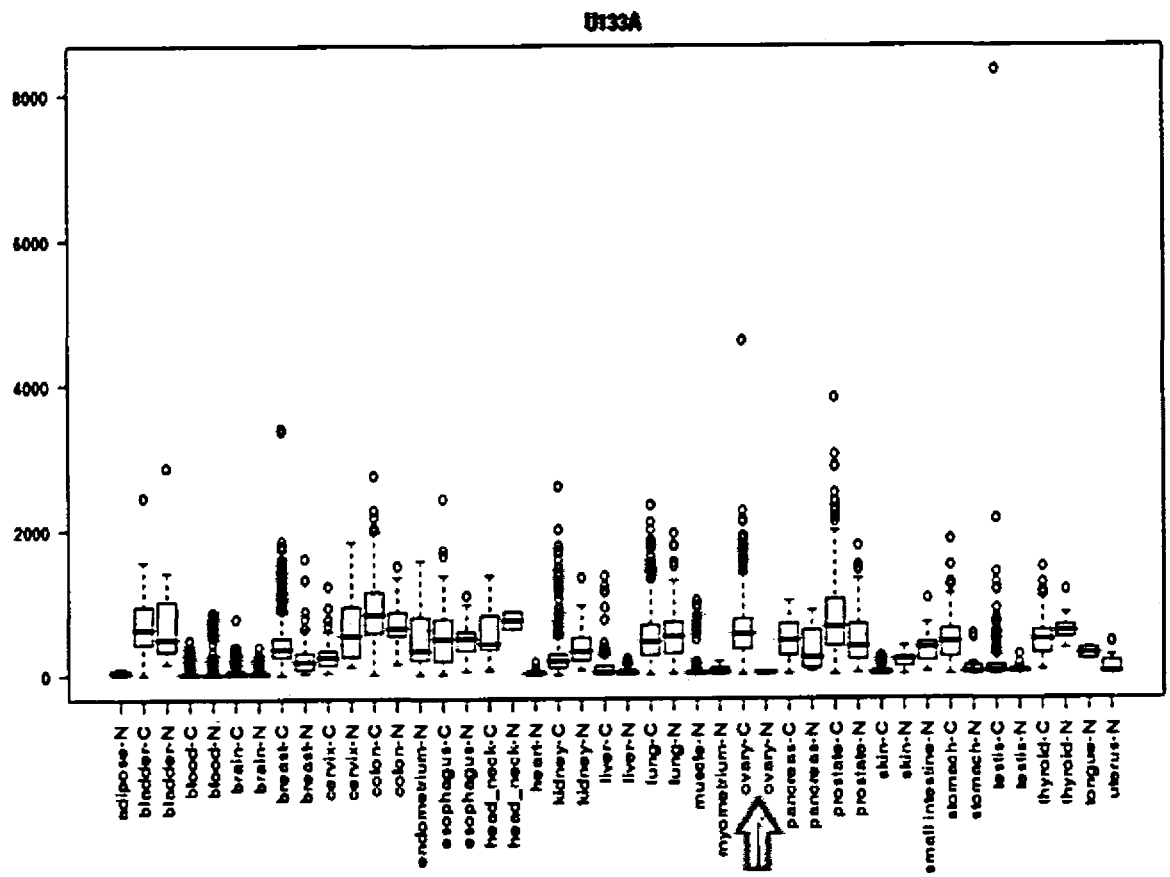
(a)


U133Plus2
10000
8000
6000
4000
2000
0
GIST-C
abdomin-C
adipose-C
adipose-N
adrenalgland-C
adrenalgland-N
bladder-C
bladder-N
blood-C
blood-N
brain-C
brain-N
breast-C
breast-N
cervix-C
cervix-N
colon-C
colon-N
endometrium-C
endometrium-N
esophagus-C
esophagus-N
head_neck-C
head_neck-N
kidney-C
kidney-N
liver-C
liver-N
lung-C
lung-N
muscle-N
ovary-C
ovary-N
pancreas-C
pancreas-N
prostate-C
prostate-N
sarcoma
sarcoma-C
skin-C
skin-N
small intestine-C
small intestine-N
stomach-C
stomach-N
testis-C
testis-N
thyroid-C
thyroid-N
tongue-N
uterus-C
uterus-N
vagina-C
vagina-N
vulva-C
vulva-N

(b)


U133Plus2
log2
0
2
4
6
8
10
12
GIST-C
abdomin-C
adipose-C
adipose-N
adrenal gland-C
adrenal gland-N
bladder-C
bladder-N
blood-C
blood-N
brain-C
brain-N
breast-C
breast-N
cervix-C
cervix-N
colon-C
colon-N
endometrium-C
endometrium-N
esophagus-C
esophagus-N
head_neck-C
head_neck-N
kidney-C
kidney-N
liver-C
liver-N
liver-T
lung-C
lung-N
muscle-N
ovary-C
ovary-N
pancreas-C
pancreas-N
prostate-C
prostate-N
sarcoma
sarcoma-C
skin-C
skin-N
small intestine-C
small intestine-N
stomach-C
stomach-N
testis-C
testis-N
thyroid-C
thyroid-N
tongue-N
uterus-C
uterus-N
vagina-C
vagina-N
vulva-C
vulva-N

(c)


U133A
8000
6000
4000
2000
0
adipose-N
bladder-C
bladder-N
blood-C
blood-N
brain-C
brain-N
breast-C
breast-N
cervix-C
cervix-N
colon-C
colon-N
endometrium-N
esophagus-C
esophagus-N
head_neck-C
head_neck-N
heart-N
kidney-C
kidney-N
liver-C
liver-N
lung-C
lung-N
muscle-N
myometrium-N
ovary-C
ovary-N
pancreas-C
pancreas-N
prostate-C
prostate-N
skin-C
skin-N
small intestine-N
stomach-C
stomach-N
testis-C
testis-N
thyroid-C
thyroid-N
tongue-N
uterus-N

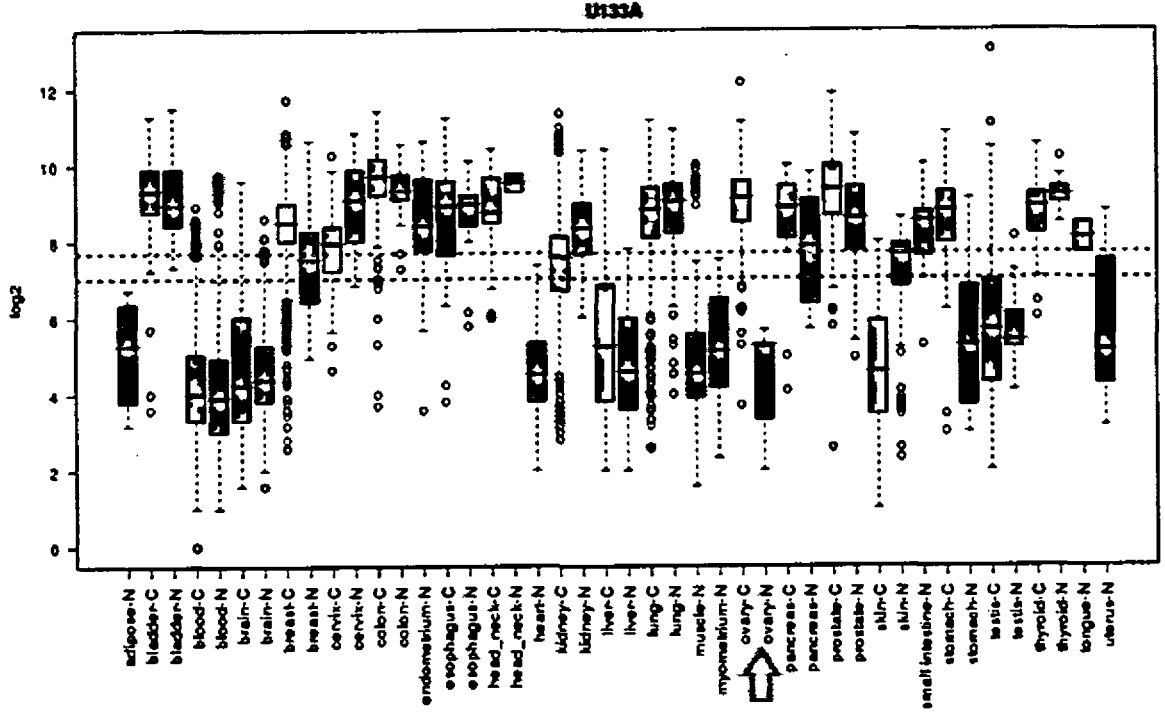
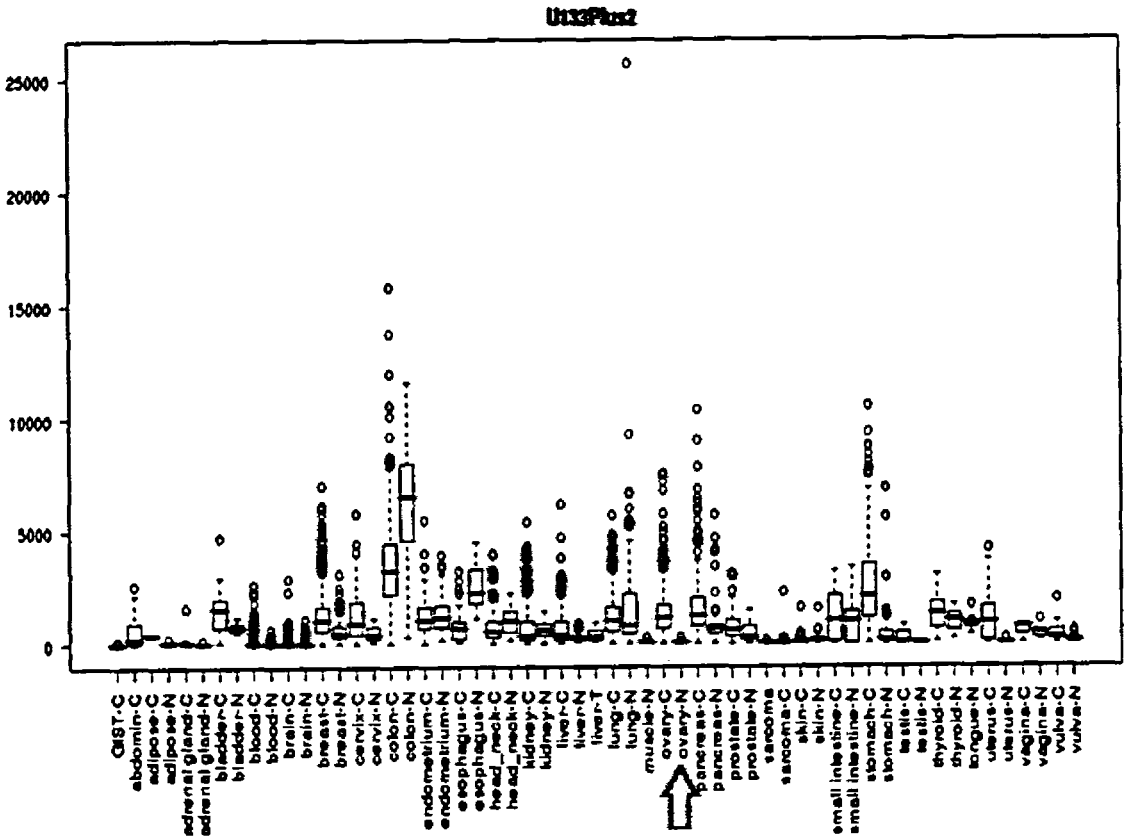


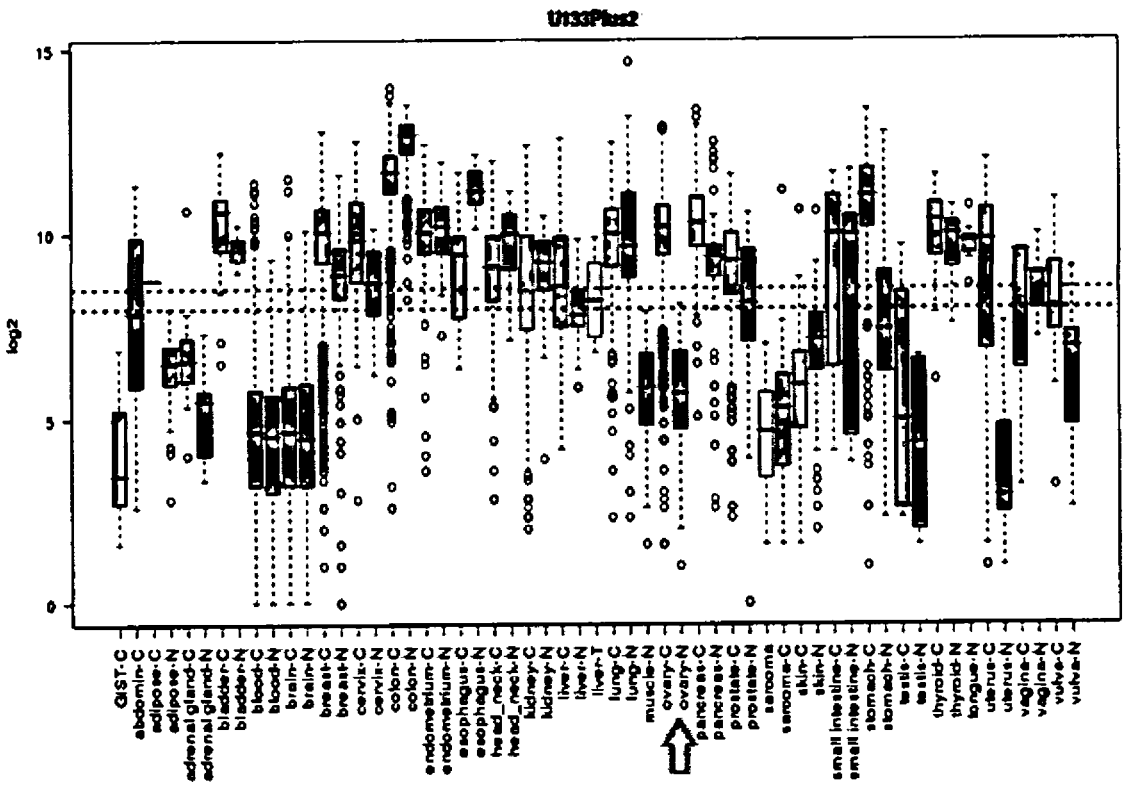
Figure 3.23: Showing up regulation of Cldn4 (A) & (B) Data set U133A (C) & (D) Data set U133A

(Ovary-C = cancerous ovary, Ovary-N = normal ovary)

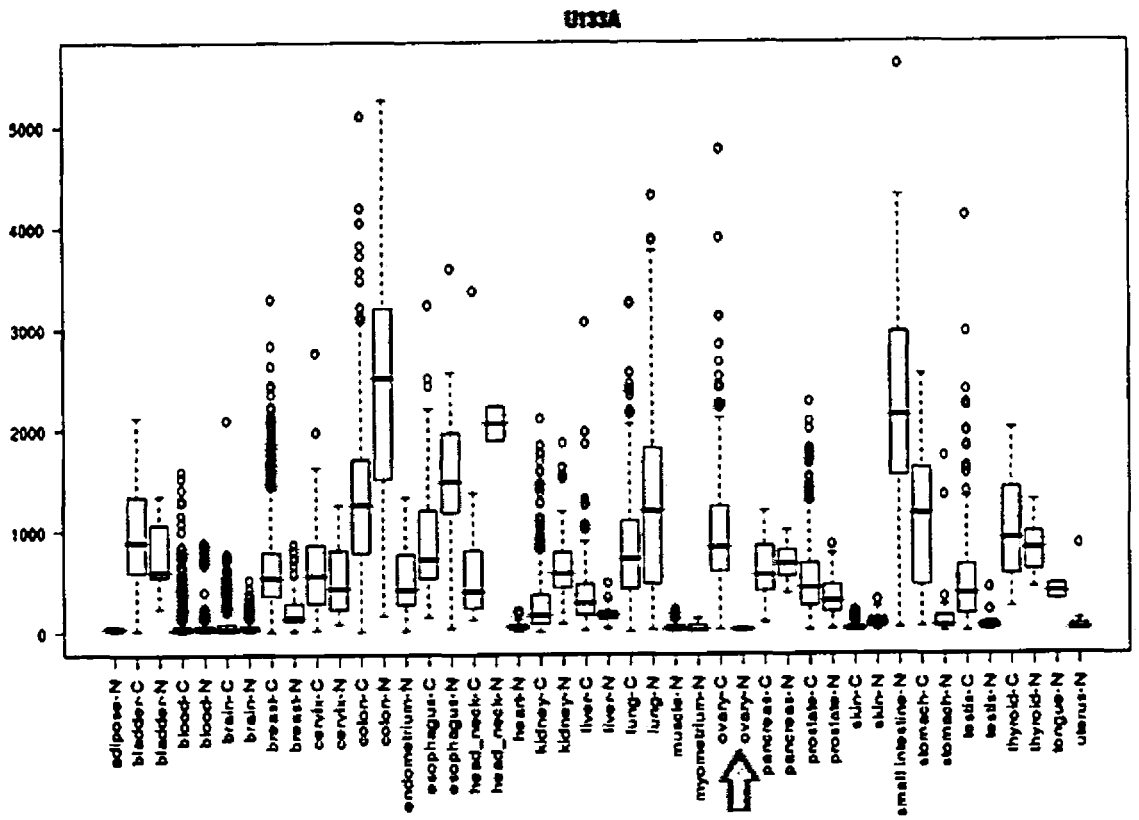
(a)


U133Plus2
25000
20000
15000
10000
5000
0
CRST-C
abdomin-C
adipose-C
adipose-N
adrenalgland-C
adrenalgland-N
bladder-C
bladder-N
blood-C
blood-N
brain-C
brain-N
breast-C
breast-N
cervix-C
cervix-N
colon-C
colon-N
endometrium-C
endometrium-N
esophagus-C
esophagus-N
head_neck-C
head_neck-N
kidney-C
kidney-N
liver-C
liver-N
liver-T
lung-C
lung-N
muscle-N
ovary-C
ovary-N
pancreas-C
pancreas-N
prostate-C
prostate-N
sarcoma
sarcoma-C
skin-C
skin-N
small intestine-C
small intestine-N
stomach-C
stomach-N
testis-C
testis-N
thyroid-C
thyroid-N
tongue-N
uterus-C
uterus-N
vagina-C
vagina-N
vulva-C
vulva-N

(b)


U133Plus2
log2
15
10
5
0
GIST-C
abdomin-C
adipose-C
adipose-N
adrenal gland-C
adrenal gland-N
bladder-C
bladder-N
blood-C
blood-N
brain-C
brain-N
breast-C
breast-N
cervix-C
cervix-N
colon-C
colon-N
endometrium-C
endometrium-N
esophagus-C
esophagus-N
head_neck-C
head_neck-N
kidney-C
kidney-N
liver-C
liver-N
liver-T
lung-C
lung-N
muscle-N
ovary-C
ovary-N
pancreas-C
pancreas-N
prostate-C
prostate-N
sarcoma
sarcoma-C
skin-C
skin-N
small intestine-C
small intestine-N
stomach-C
stomach-N
testis-C
testis-N
thyroid-C
thyroid-N
tongue-N
uterus-C
uterus-N
vagina-C
vagina-N
vulva-C
vulva-N

(c)


U133A
5000
4000
3000
2000
1000
0
adipose-N
bladder-C
bladder-N
blood-C
blood-N
brain-C
brain-N
breast-C
breast-N
cervix-C
cervix-N
colon-C
colon-N
endometrium-N
esophagus-C
esophagus-N
head_neck-C
head_neck-N
heart-N
kidney-C
kidney-N
liver-C
liver-N
lung-C
lung-N
muscle-N
myometrium-N
ovary-C
ovary-N
pancreas-C
pancreas-N
prostate-C
prostate-N
skin-C
skin-N
small intestine-N
stomach-C
stomach-N
testis-C
testis-N
thyroid-C
thyroid-N
tongue-N
uterus-N

(d)

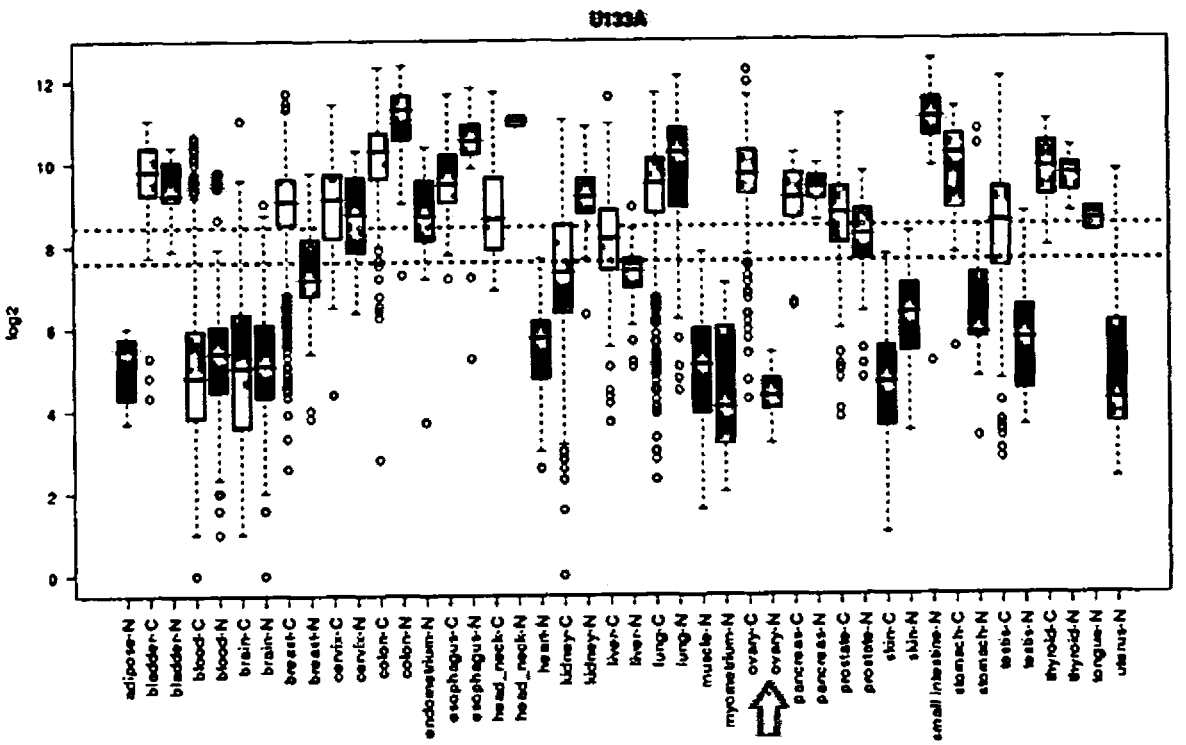
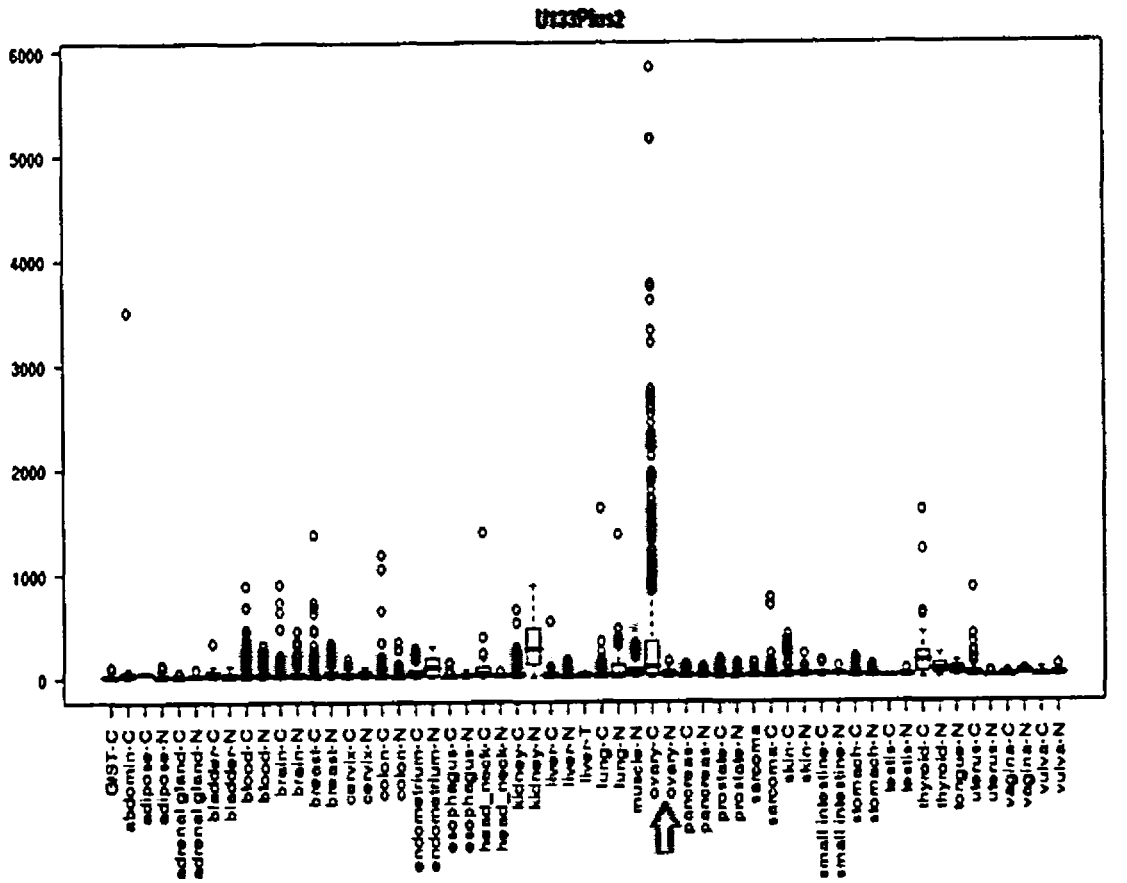


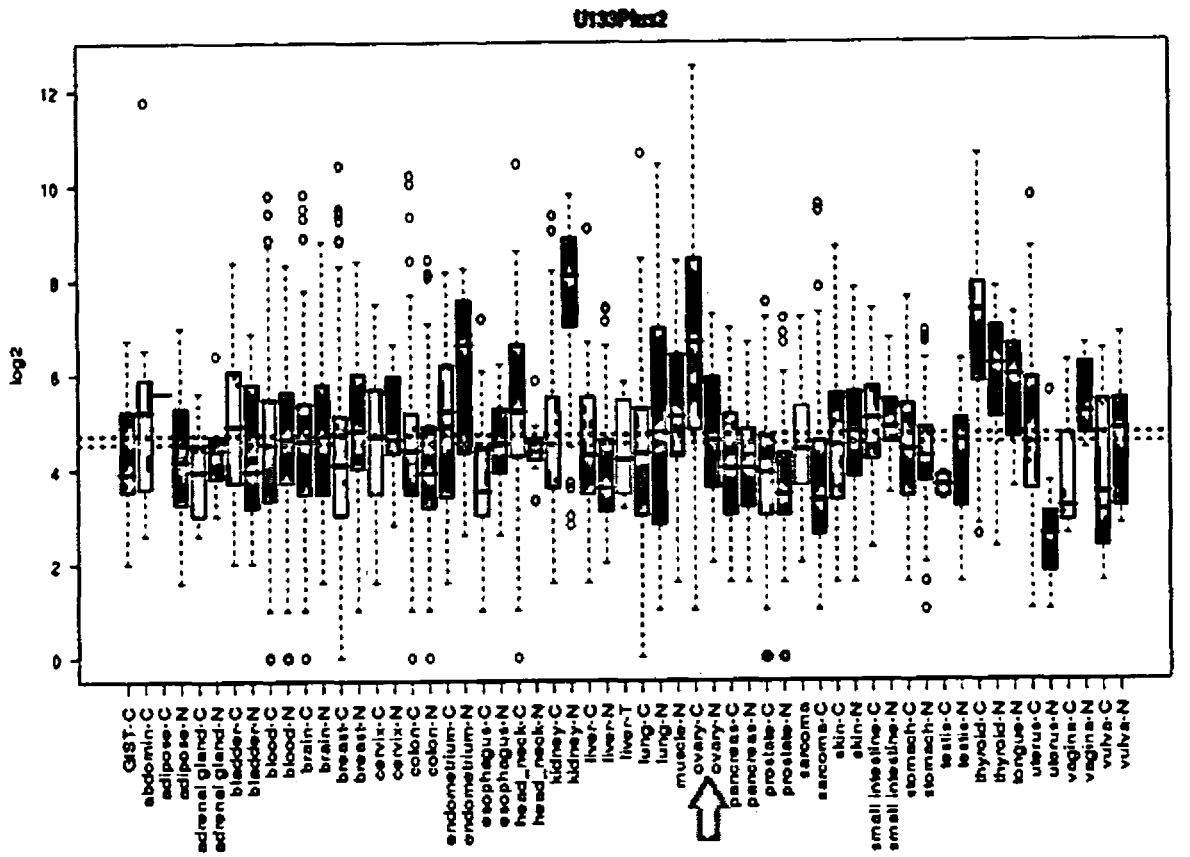
Figure 3.24: Showing up regulation of Cldn7 (A) & (B) Data set U133A (C) & (D) Data set U133A

(Ovary-C = cancerous ovary, Ovary-N = normal ovary)

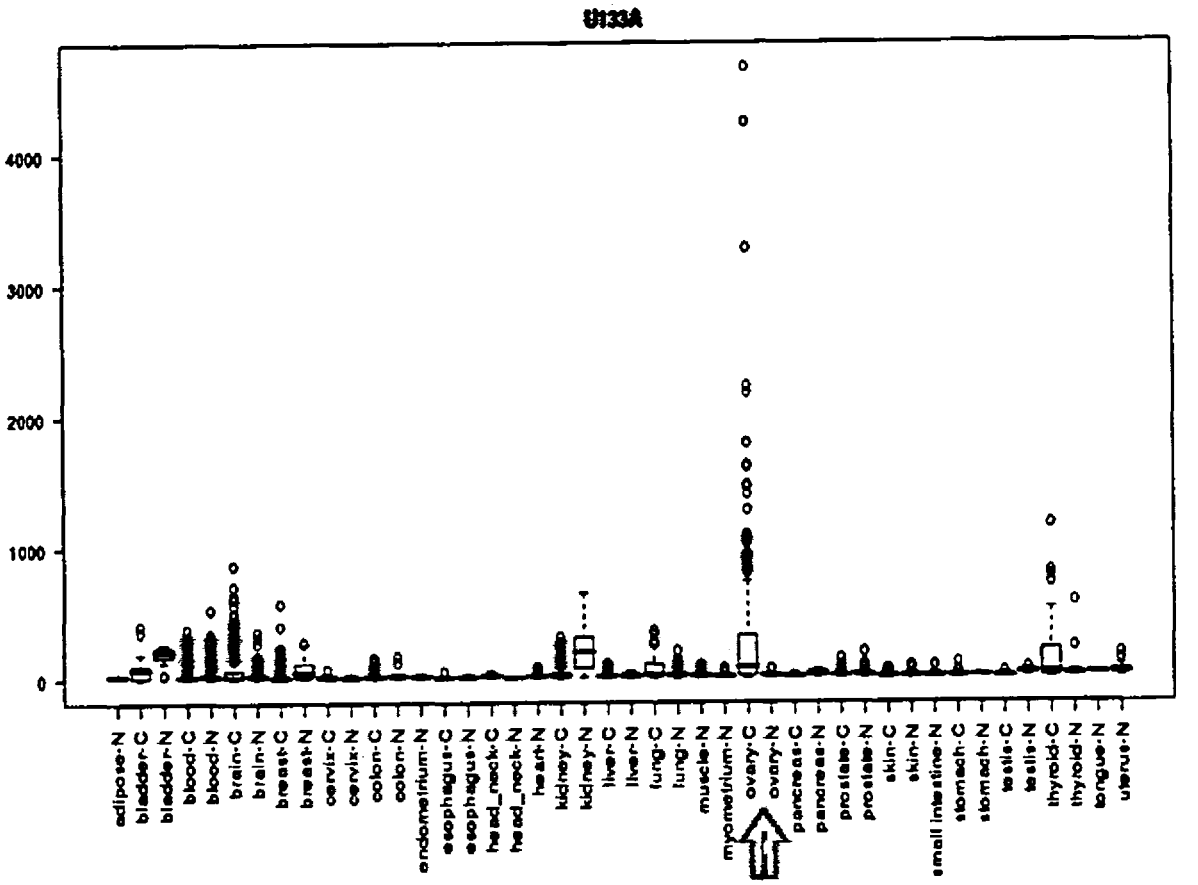
(a)


U133Plus2
6000
5000
4000
3000
2000
1000
0
GNF.C
abdomin.C
adipose.C
adipose.
adrenal gland.C
adrenal gland.
bladder.C
bladder.
blood.C
blood.
brain.C
brain.
breast.C
breast.
cervix.C
cervix.
colon.C
colon.
endometrium.C
endometrium.
esophagus.C
esophagus.
head_neck.C
head_neck.
kidney.C
kidney.
liver.C
liver.
lung.C
lung.
muscle.C
ovary.C
ovary.
pancreas.C
pancreas.
prostate.C
prostate.
sarcoma.C
sarcoma.
skin.C
skin.
small intestine.C
small intestine.
stomach.C
stomach.
testis.C
testis.
thyroid.C
thyroid.
tongue.C
tongue.
uterus.C
uterus.
vagina.C
vagina.
vulva.C
vulva.

(b)


U133Plus2
log2
0
2
4
6
8
10
12
GST-C
abdomin-C
adipose-C
adipose-N
adrenal gland-C
adrenal gland-N
bladder-C
bladder-N
blood-C
blood-N
brain-C
brain-N
breast-C
breast-N
cervix-C
cervix-N
colon-C
colon-N
endometrium-C
endometrium-N
esophagus-C
esophagus-N
head_neck-C
head_neck-N
kidney-C
kidney-N
liver-C
liver-N
liver-T
lung-C
lung-N
muscle-N
ovary-C
ovary-N
pancreas-C
pancreas-N
prostate-C
prostate-N
sarcoma
sarcoma-C
skin-C
skin-N
small intestine-C
small intestine-N
stomach-C
stomach-N
testis-C
testis-N
thyroid-C
thyroid-N
tongue-N
uterus-C
uterus-N
vagina-C
vagina-N
vulva-C
vulva-N

(c)


U133A
4000
3000
2000
1000
0
adipose-N
bladder-C
bladder-N
blood-C
blood-N
brain-C
brain-N
breast-C
breast-N
cervix-C
cervix-N
colon-C
colon-N
endometrium-N
esophagus-C
esophagus-N
head_neck-C
head_neck-N
heart-N
kidney-C
kidney-N
liver-C
liver-N
lung-C
lung-N
muscle-N
myometrium-N
ovary-C
ovary-N
pancreas-C
pancreas-N
prostate-C
prostate-N
skin-C
skin-N
small intestine-N
stomach-C
stomach-N
testis-C
testis-N
thyroid-C
thyroid-N
tongue-N
uterus-N

(d)

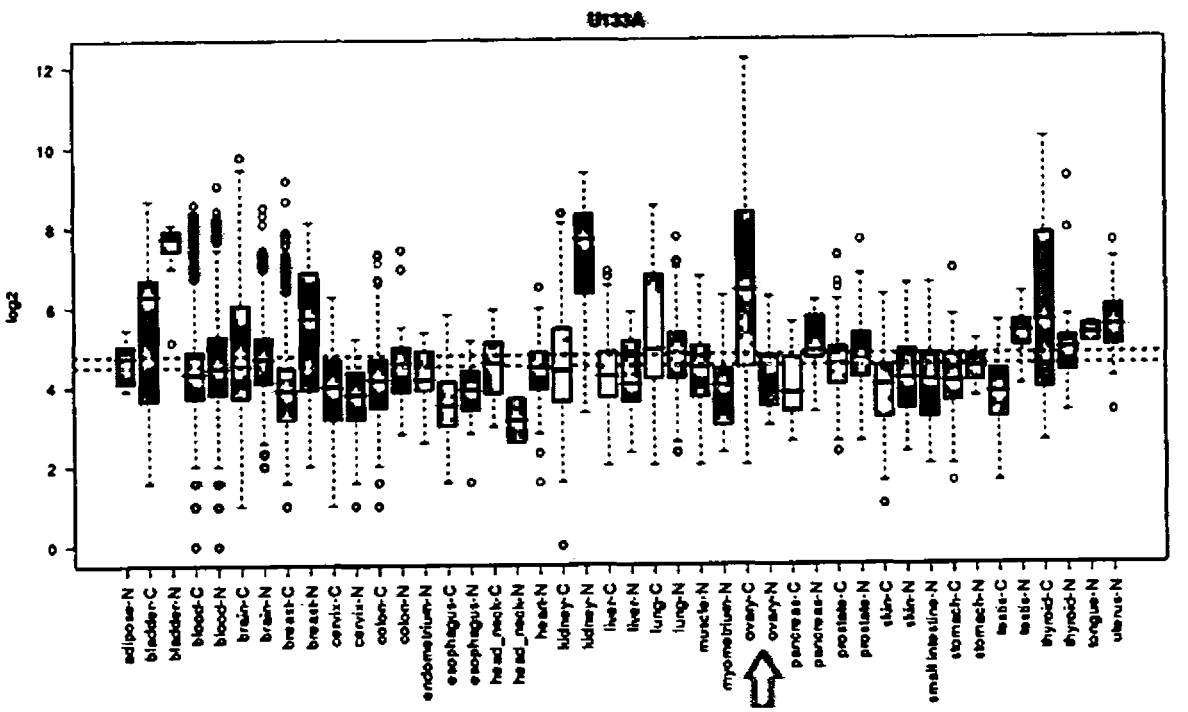
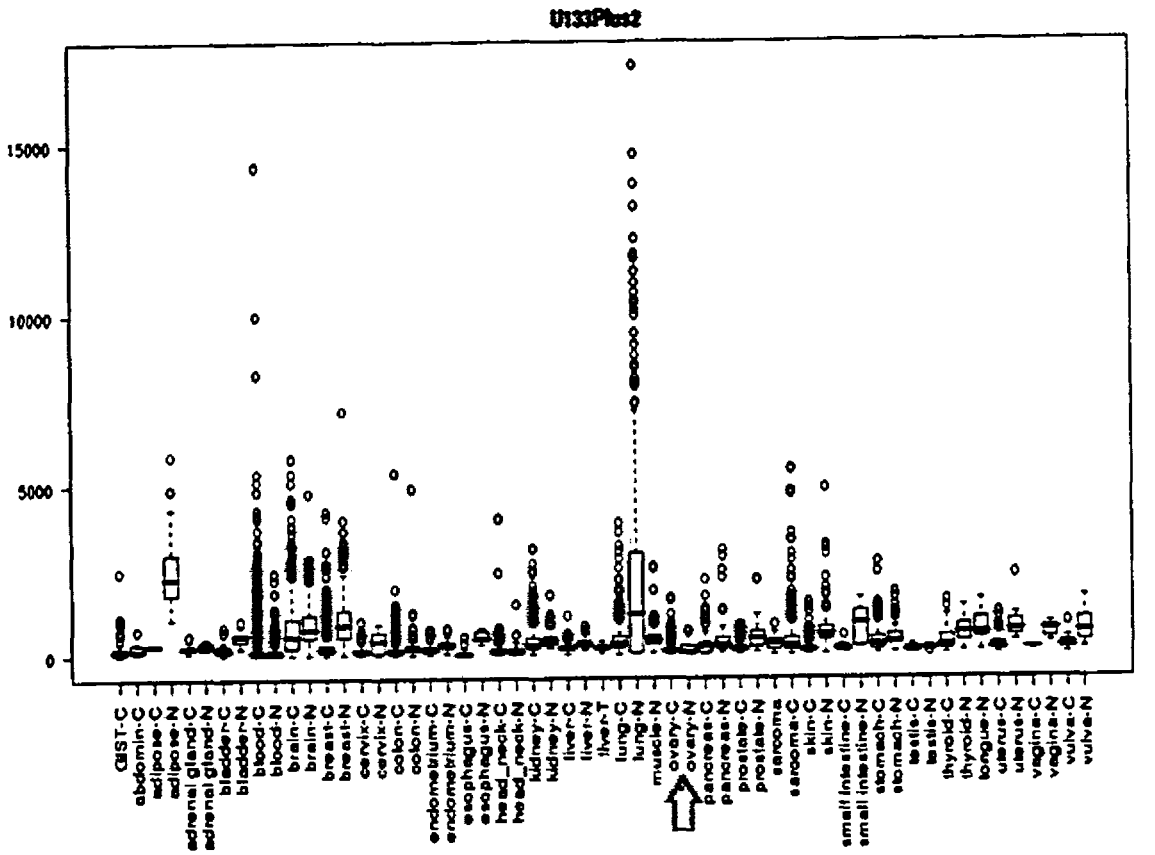


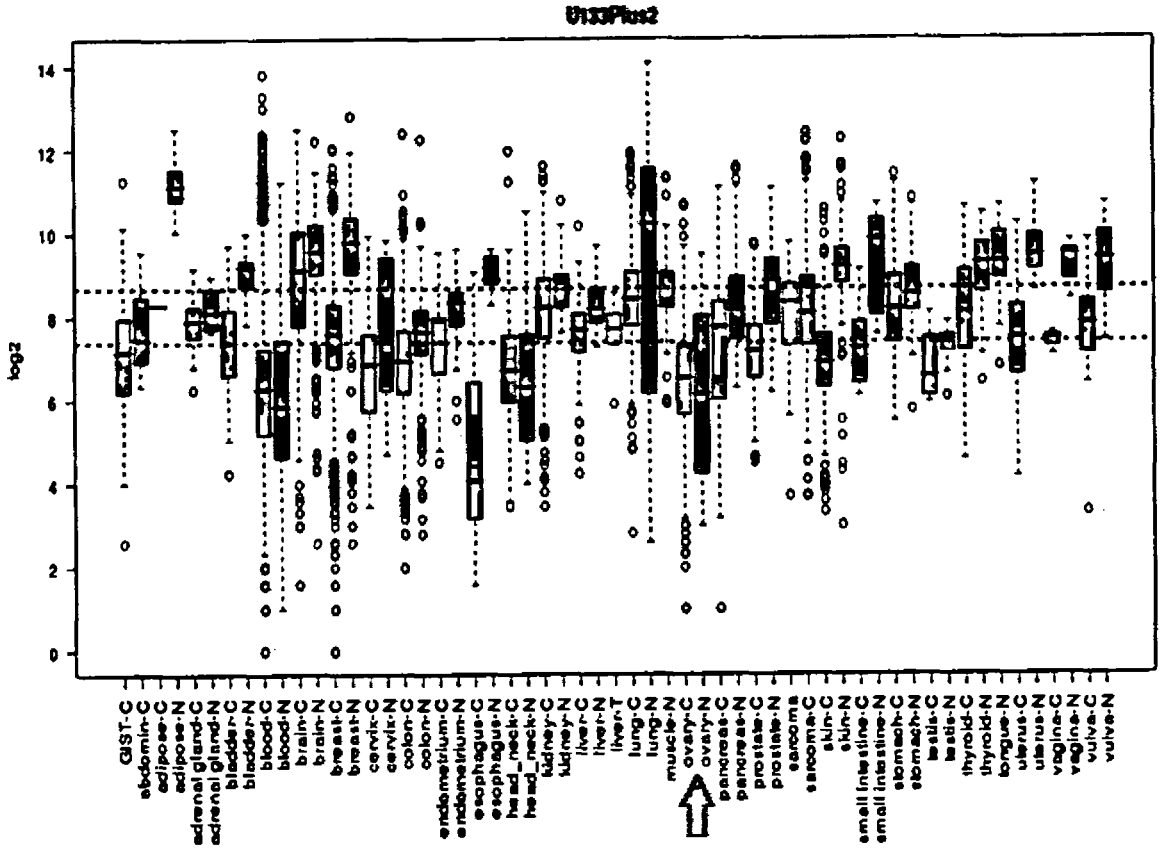
Figure 3.25: Showing up regulation of Cldn16 (A) & (B) Data set U133A (C) & (D) Data set U133A

(Ovary-C = cancerous ovary, Ovary-N = normal ovary)

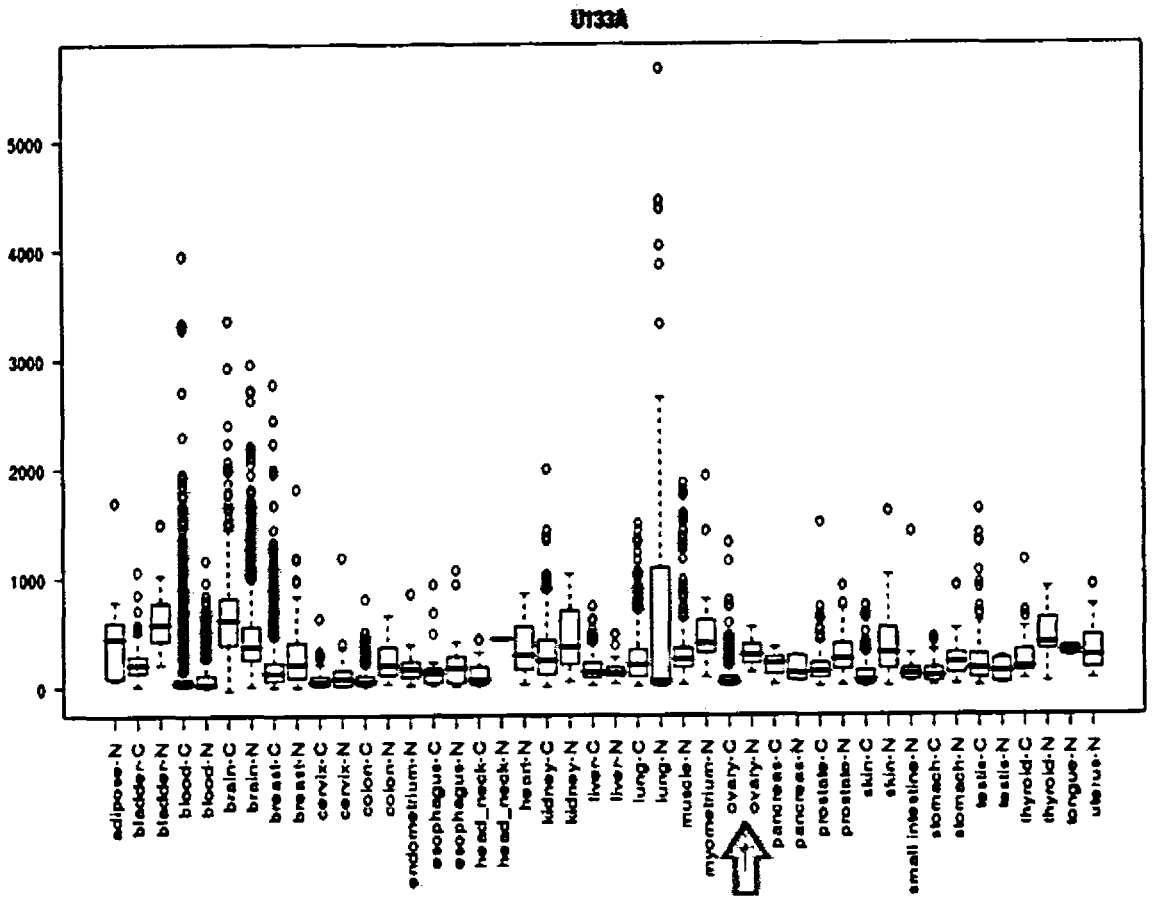
(a)


U133Plus2
15000
10000
5000
0
GIST-C
abdomin-C
adipose-C
adipose-N
adrenal gland-C
adrenal gland-N
bladder-C
bladder-N
blood-C
blood-N
brain-C
brain-N
breast-C
breast-N
cervix-C
cervix-N
colon-C
colon-N
endometrium-C
endometrium-N
esophagus-C
esophagus-N
head_neck-C
head_neck-N
kidney-C
kidney-N
liver-C
liver-N
liver-T
lung-C
lung-N
muscle-N
ovary-C
ovary-N
pancreas-C
pancreas-N
prostate-C
prostate-N
sarcoma
sarcoma-C
skin-C
skin-N
small intestine-C
small intestine-N
stomach-C
stomach-N
testis-C
testis-N
thyroid-C
thyroid-N
tongue-N
uterus-C
uterus-N
vagina-C
vagina-N
vulva-C
vulva-N

(b)


U133Plus2
log2
0
2
4
6
8
10
12
14
GIST-C
abdomin-C
adipose-C
adipose-N
adrenal gland-C
adrenal gland-N
bladder-C
bladder-N
blood-C
blood-N
brain-C
brain-N
breast-C
breast-N
cervix-C
cervix-N
colon-C
colon-N
endometrium-C
endometrium-N
esophagus-C
esophagus-N
head_neck-C
head_neck-N
kidney-C
kidney-N
liver-C
liver-N
lung-C
lung-N
muscle-N
ovary-C
ovary-N
pancreas-C
pancreas-N
prostate-C
prostate-N
sarcoma
sarcoma-C
skin-C
skin-N
small intestine-C
small intestine-N
stomach-C
stomach-N
testis-C
testis-N
thyroid-C
thyroid-N
tongue-C
uterus-C
uterus-N
vagina-C
vagina-N
vulva-C
vulva-N

(c)


U133A
5000
4000
3000
2000
1000
0
adipose-N
bladder-C
bladder-N
blood-C
blood-N
brain-C
brain-N
breast-C
breast-N
cervix-C
cervix-N
colon-C
colon-N
endometrium-N
esophagus-C
esophagus-N
head_neck-C
head_neck-N
heart-N
kidney-C
kidney-N
liver-C
liver-N
lung-C
lung-N
muscle-N
myometrium-N
ovary-C
ovary-N
pancreas-C
pancreas-N
prostate-C
prostate-N
skin-C
skin-N
small intestine-N
stomach-C
stomach-N
testis-C
testis-N
thyroid-C
thyroid-N
tongue-N
uterus-N

(d)

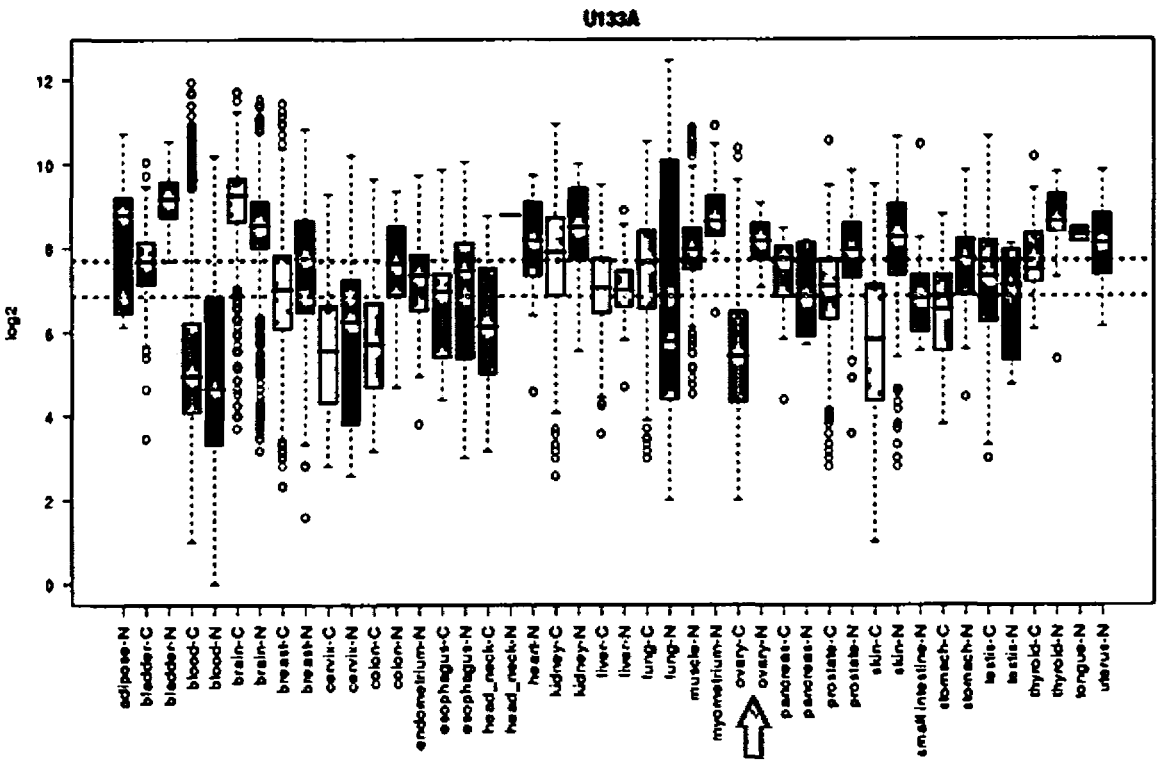


Figure 3.26: Showing up regulation of Cldn5 (A) & (B) Data set U133A (C) & (D) Data set U133A.

(Ovary-C = cancerous ovary, Ovary-N = normal ovary)

Table 3.5 Microarray claudin family results of dataset GSE6008 & GSE4122

Claudin genes	**Reference ID**	**Fold change (GSE6008)**	**Fold change (GSE4122)**	**P-value**	**False discovery rate**
cldn4	201428_at	> 2	> 5	0.0018636247	0.003261343
cldn7	202790_at	> 2	> 3	2.9067648E-6	6.7824512E-6
cldn3	203953_s_at	> 3	> 6	0.0	0.0
cldn5	204482_at	< 2	< 3	0.009433004	0.011005172
cldn15	219640_at	no change	no change	1.7732685E-8	6.2064395E-8
cldn16	220332_at	> 1.5	> 3	0.004344086	0.0060817203

Table 3.6 SAGE results for claudin family

Gene	Unigene Cluster	SAGE Tag	Normal (tpm)		Ovarian Cancer (tpm)	
			Tissues	Cell lines	Tissues	Cell lines
Cldn3	Hs.647023	CTCGCGCTGG	No data	0	77	0
Cldn4	Hs.647036	ATCGTGGCGG	No data	0	91	0
Cldn7	Hs.513915	TATAGTCCTC	No data	0	37	0
Cldn16	Hs.251391	TTGCCATCCT	No data	0	0	0
Cldn5	Hs.505337	GACCGCGGCT	No data	0	14	0

Table 3.7 Selected MicroRNA's for claudin genes with minimum energies

Gene ID	MicroRNA	Chromosomal location	Energy
Cldn3	hsa-miR-423	17q11.2	-86.3 kcal/mol
	hsa-miR-10b	2q31.1	-76.7 kcal/mol
Cldn4	hsa-miR-572	4p15.33	-119.5 kcal/mol
	hsa-miR-149	2q37.3	-109.9 kcal/mol
	hsa-miR-612	11q13.1	-105.9 kcal/mol
Cldn7	hsa-miR-339	7p22.3	-117.5 kcal/mol
	hsa-miR-506	Xq27.3	-97.0 kcal/mol
	hsa-miR-642	19q13.32	-93.5 kcal/mol
Cldn16	hsa-let-7e	19q13.41	-70.2 kcal/mol
	hsa-miR-224	Xq28	-65.4 kcal/mol

Chapter No 04

Discussion

4. DISCUSSION

The participation of microRNA's in cancers is a major breakthrough in cancer therapeutic research. The microRNA's can be best described as a regulator of their targeted genes (Lai, 2003; Bartel, 2004). This process of regulation is at post transcriptional level of the gene. They bind to the complementary sites present on the genes to which they target and hence block the translation and results in truncated products. The discovery of microRNA's forced the cancer scientists to think that only examining the expression of genes in cancer state is not sufficient but along with it, it is necessary to explore those genes which are actually regulating their expression. Hence the microRNA's are behind the scene players of the story of gene expression and without their presence whole phenomena of gene expression can never be well understood.

The role of microRNA in cancer can never be denied. But the major problem which arises is that the accurate targets for all of the identified microRNA's are not known yet. So it is very difficult to check the microRNA's involved with cancers. Although some of the wet lab analysis techniques had been proposed for the solution of this problem but these techniques are very expensive and difficult. One way to solve this problem is to design a computer based technique which should be easier, accurate and inexpensive. In current research a unique method is proposed to identify microRNA's involved in human ovarian cancer. The method is easiest and unique in the way that it indirectly identifies some of the microRNA's potentially involved in ovarian cancer. As it is known that expression of many genes has been regulated by microRNA's so if the down or up regulation of that gene is involved in appearance of any disease then it can be

said that those microRNA's which are in fact regulating this gene are indirectly involved in the expression of this gene. This hypothesis is the basis of current research. Many times in case of cancers the down/up regulation of gene or the mutations in the gene are not only the reason, many times the problem lies in microRNA's which are targeting them and regulating their expression. So this suggests that if it is known that a gene is involved in a specific cancer and it is also known that this gene is targeted by which microRNA then microRNA's which can be potential targets for the treatment of that disease can be easily sorted out. For the absolute understanding of disease pathogenecity the gene expression analysis must be done inclusive with microRNA gene expression. The current research is the example of such investigation. This method is straightforward and is of two steps. First, is to extract genes associated with the human ovarian cancer and second, is to identify the potential microRNA's regulating those genes. Both steps were designed by adopting computational approaches.

The microarray analysis of ovarian cancer data sets resulted in a collection of genes that can be the potential targets for disease. Two gene families named as pyruvate dehydrogenase kinase family and claudin family, found to have major role in development of ovarian cancer. Pathogenesis of both gene families in ovarian cancer and their putative microRNA's will be discussed individually.

The results revealed PDK2 and PDK4 as novel biological markers for ovarian cancer. PDK's play a potential role in energy metabolism of the cell and any mutation in them can cause cancer (Mayevsky, 2009; Seyfried and Mukherjee, 2005; Chen *et al.*, 2009; Ramanathan *et al.*, 2005). To date the role of PDK's is unassigned in ovarian

cancer. The PDK's major role is in phosphorylation which in turn makes them more important in cases of diabetes and starvation (Jeoung and Harris, 2010). Naturally in case of starvation or fatty diet, PDK's are up regulated and this up regulation shut down the PDC activity which in turn preserves the pyruvate, lactate, and alanine (Majer *et al.*, 1998; Guerre-Millo *et al.*, 2000). These carbon compounds are involved in glucose production and hence their conservation reduces the deposition of fats. This preservation of carbon compounds is very crucial for the survival in critical conditions of starvation and this job is done by PDK's. Now how this phenomenon of PDK activity interacts in case of cancer is a very tricky question. While answering this query it was realized that the biological functioning of PDK's was never fully understood.

The analysis of ovarian cancer expression data showed the significant over expression of PDK2 and PDK4, which disclosed their activity in cancer. As tumors have high energy demands and so in order to fulfill this requirement there is need of continuous glucose production and for this the conversion of carbon compounds in to fats should be blocked. This job can be performed by down regulating the PDC activity by the help of advanced regulation of PDK4 and PDK2. The current research verify their over expression in ovarian cancer but normally there is no or very little expression of these PDK's. So the enhanced level of expression in cancer supports the assumption about the process for the fulfillment of high energy demands of tumor cells (Wende *et al.*, 2005; Araki and Motojima, 2006; Zhang *et al.*, 2006).

Another hypothesis for the up regulation of these PDKs in cancer is that they are involved in the enhancement of the resistance mechanism of tumors and make them able

to withstand the hypoxic conditions produced as a result of greater glucose uptake of tumors (Wende *et al.*, 2005; Araki and Motojima, 2006; Zhang *et al.*, 2006). So for tumor growth and development their expression level must be enhanced, apart from their normal behavior. The role of PDK4 is little strange as it shows down regulation in tissues while performing SAGE and also in GENT one data set shows down regulation but at the same time it was noticed that this down regulation is very minimal so it could be neglected because PDK4 is also over expressed in cell lines. As there are strong evidences of over expression of PDK4 in ovarian cancer so this minimal down expression was ignored.

The microRNA's can be best described as a regulator of their targeted genes. This process of regulation is at posttranscriptional level of the gene. They bind to the complementary sites present on the genes to which they target and hence block the translation and results in truncated products. The discovery of microRNA's forced the cancer scientists to think that only examining the expression of genes in cancer state is not sufficient but along with it, it is necessary to explore those genes which are actually regulating their expression. Hence the microRNA's are behind the scene players of the story of gene expression and without their presence whole phenomena of gene expression can never be well understood. In current research after finding the PDK2 and PDK4 role in ovarian cancer their potential targeting microRNA's was also predicted by different reliable and robust prediction algorithms. The increased differential expression of PDK2 and PDK4 in ovarian cancer reveals that their targeting microRNA's expression must be lowered during cancer state and therefore the over expression of these genes is possible.

Because if the microRNA's responsible for the control of PDK2 and PDK4 expression, become higher in percentage then usual then their blocking ability will be enhanced which will result in little or no expression of these genes in ovarian cancer. The microarray analysis of these genes strongly convince about their over enhanced role in ovarian cancer so it was thought that all of the microRNA's for these genes in case of ovarian cancer must be down regulated.

Recent studies also described that in case of cancers the microRNA's mostly happen to be down regulated (Meng *et al.*, 2006; Lu *et al.*, 2005). Some of the predicted microRNA's were never investigated before for their involvement in ovarian cancer. Scientists always linked the ovarian cancer with the breast cancer on the basis of their almost similar genetics (Bergfeldt *et al.*, 2002). We predicted hsa-miR-497 as regulator of PDK4 in ovarian cancer. In a very recent study of February 2011, Li *et al.* reported the downregulation of has-miR-497 in breast cancer (Li *et al.*, 2011). Similarly the expression of hsa-miR-326 was also reported in breast cancer (Liang Z *et al.*, 2010). These studies made these microRNA's as a strong candidate for their involvement in ovarian cancer. hsa-miR-424, hsa-miR-195 and hsa-miR-15a were already reported to be down regulated in case of ovarian cancer (Dahiya *et al.*, 2008; Zhang *et al.*, 2008). While role of other predicted microRNA's hsa-miR-1208, hsa-miR-330-5p, hsa-miR-1224-5p is ambiguous but future research on these microRNA's will untangle many interesting features as potential biological markers of ovarian cancer.

The second identified gene family is the claudin family. Claudins are tight junction proteins and their involvement in several cancers implicated their role in tumor

development. Some of the claudin family members association with the ovarian cancer has already been identified in vitro but this is the first in silico analysis of complete claudin gene family's association specifically with the ovarian cancer. As it is described earlier that three members of claudin family has already been reported for their involvement in expression of ovarian cancer, results not only verified the significant up regulation of these genes but the over expression of cldn16 was also observed in cancer state. Cldn16 showed up regulation in both SAGE and Microarray analysis. Although the expression of cldn16 is less than the cldn4, cldn3 and cldn7 as it is clear in volcano plot and SAM graph but its involvement in cancer can never be neglected. The clustering analysis also confirmed this finding although the percentage of up regulation is not so much high. GENT gave a very interesting result by showing highest percentage of up regulated cldn16 in cancerous ovary. This confirmed the over expression of cldn16 in human ovarian cancer.

Another interesting finding was the down regulation of cldn5 in microarray data sets, which is very amazing as far as the role of claudins is observed. The tight junction formation ability of claudins makes to believe their up regulating role in tumor formation but here the association of cldn5 as significant down regulated gene in ovarian cancer is the most surprising thing which reveals the fact that may be the scientific knowledge about the claudins is still very limited and there are many other unraveled truths about the claudins that have to be identified yet. But the SAGE analysis of cldn5 revealed the totally different results by showing its up regulation in cancerous ovary. GENT also showed the same results as of SAGE. These findings about the cldn5 makes it role

suspicious in ovarian cancer which must be analyzed in vitro. But here it can be stated that there are more evidences of its up regulation rather than down regulation when the overall molecular functioning of claudins was analyzed. However the microarray results can not be over looked because these results are obtained from real microarray data. Similarly SAGE is also a very reliable tool because it is based upon DNA sequencing data. So the conflicting results of SAGE and microarray demands more extensive research on the role of cldn5 in ovarian cancer.

SAGE and microarray analysis revealed that cldn3, cldn4, cldn7 and cldn16 were over expressed in human ovarian cancer. So only these caludin members are picked for prediction of their corresponding microRNA's. Four prediction tools were used for that purpose. All of them predicted number of associated microRNA's. As the method for prediction of each tool is different on the basis of different criteria's so every tool predicted some microRNA's that are different from other. Only those microRNA's were selected which were in common among at least two tools. Then after that the selected microRNA's were checked for their hybridization energies. The energy with which microRNA binds with mRNA is very important for its binding with target and then its regulation. If the energy is lower, then the binding is more perfect. The greater energy accounts for poor binding and so the less involvement of that microRNA in regulation of that gene. The top three microRNA's having minimum free binding energy with each gene were then selected as potential regulators of that gene expression. The minimum energies of these microRNA's made them very suitable as most appropriate regulators of their corresponding targets. So in simple words it can be said that above mentioned all of

the microRNA's are associated with human ovarian cancer. As all of these predictions are in silico so there is a need of in vitro verification of these findings.

Another interesting observation was made in the results that some microRNA's are common for two or more genes of this family. For example hsa-miR-10a, hsa-miR-10b and hsa-miR-423 are common regulators of cldn3, cldn4 and cldn7. This means that these microRNA's specifically regulate the whole claudin gene family and also in case of ovarian cancer these microRNA's might have more probability of association.

The combined microRNA results of both families disclosed that few microRNA's occupy almost same chromosomal location or they are present on same chromosome. This represents that few chromosomal locations were very hot for finding microRNA's responsible for ovarian cancer. hsa-miR-497 and hsa-miR-195 are found on 17p13.1 while hsa-miR-423 was present on 17q11.2 which means that chromosome 17 contains the putative microRNA's for ovarian cancer. Similarly on 19q13.32 hsa-miR-642 and hsa-miR-330-5p were present and very near to them on 19q13.41 position hsa-let-7e resides. So chromosome 19 arm was very important for screening of microRNA's in ovarian cancer. Another example is of hsa-miR-506, hsa-miR-612 and hsa-miR-424 which belong to the q arm of chromosome X. hsa-miR-149 and hsa-miR-10b are from chromosome 2 arm q and hsa-miR-16 and hsa-miR-15a are on 13q14.2. The above mentioned facts describe that chromosome 2, 13, 17, 19 and X are the hottest chromosomes for finding microRNA involved in ovarian cancer. Another interesting thing is that only q arm of these chromosomes contain these microRNA's. At this stage it is very hard to understand the biological relevance of this finding but it must be further studied and may result in some new insights for microRNA cancer research. For the

ovarian cancer the mentioned chromosomes must be fully scanned again for more putative markers.

Although all computational findings can never hundred percent correct and there is always chances of errors so the results are not always accurate but the in silico methods provide the way for the solution. It is very difficult to analyze all of the microRNA's in vitro for its association with ovarian cancer but computational analysis is very helpful in short listing only those candidates which have high probability of their involvement in disease. So this study provides a method for minimizing the burden for in vitro analysis. Instead of scanning full database of microRNA's for their association with disease, researchers should only have to concentrate on most credible targets.

REFERNCES

- Ao A, Wang H, Kamarajugadda S and Lu J (2008) Involvement of estrogen-related receptors in transcriptional response to hypoxia and growth of solid tumors. *Proc Natl Acad Sci U S A,* 105(22), 7821–7826.
- Abrahante JE, Daul AL, Li M, Volk ML, Tennessen JM, Miller EA and Rougvie AE (2003) The *Caenorhabditis elegans* hunchback-like gene lin-57/hbl-1 controls developmental time and is regulated by microRNAs, *Dev. Cell,* **4,** 625–637.
- Anderson JM and Itallie CM (1995) Tight junctions and the molecular basis for regulation of paracellular permeability, *Am J Physiol,* 269, G467– G475.
- Araki M and Motojima K (2006), Identification of ERRalpha as a specific partner of PGC-1alpha for the activation of PDK4 gene expression in muscle, *FEBS J,* 273, 1669–1680.
- Aukerman MJ and Sakai H (2003) Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes, *Plant Cell,* 15, 2730–2741.
- Babak T, Zhang W, Morris Q, Blencowe BJ and Hughes TR (2004) Probing microRNAs with microarrays: Tissue specificity and functional inference, *RNA,* 10, 1813-1819.
- Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function, *Cell,* 116(2), 281-97.
- Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau W-C, Ledoux P, Rudnev D, Lash AE, Fujibuchi W and Edgar R (2005) NCBI GEO: mining millions of

expression profiles—database and tools, *Nucl. Acids Res,* 33 *(suppl 1): D562-D566.*

- Barrett T and Edgar R (2006) Gene Expression Omnibus (GEO): Microarray data storage, submission, retrieval, and analysis, *Methods Enzymol*, 411, 352-369.
- Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Tomashevsky M and Edgar R (2007) NCBI GEO: mining tens of millions of expression profiles—database and tools update, *Nucl. Acids Res*, 35, D760-D765.
- Bergfeldt K, Rydh B, Granath F, Grönberg H, Thalib L, Adami H-O and Hall P (2002) Risk of ovarian cancer in breast-cancer patients with a family history of breast or ovarian cancer: a population-based cohort study, *The Lancet*, 360 (9337), 891 - 894.
- Betel D, Koppal A, Agius P, Sander C, Leslie C, (2010) Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites, *Genome Biology*, 11, R90.
- Betel D, Wilson M, Gabow A, Marks DS and Sander C (2008) The microRNA.org resource: targets and expression, *Nucleic Acids Res*, 36, D149-53.
- Bignotti E, Tassi RA, Calza S, Ravaggi A, Romani C, Rossi E, Falchetti M, Odicino FE, Pecorelli S and Santin AD (2006) Differential gene expression profiles between tumor biopsies and short-term primary cultures of ovarian serous carcinomas: Identification of novel molecular biomarkers for early diagnosis and therapy, *Gynecologic Oncology*, 103 (2), 405-416.

- Bowker-Kinley MM, Davis WI, WU P, Harris RA and Popov KM (1998) Evidence for existence of tissue-specific regulation of the mammalian pyruvate dehydrogenase complex, *Biochem. J,* 329, 191-196.
- Boylan KLM., Misemer B, DeRycke MS, Andersen JD, Harrington KM, Kalloger SE, Gilks CB, Pambuccian SE and Skubitz APN (2011) Claudin 4 is differentially expressed between ovarian cancer subtypes and plays a role in spheroid formation, *International Journal of Molecular Sciences,* 12, 1334-1358.
- Brennecke J, Hipfner DR, Stark A, Russell RB and Cohen SM (2003) Bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in *Drosophila*, *Cell,* 11, 25–36.
- Cai X, Schäfer A, Lu S, Bilello JP, Desrosiers RC, Edwards R, Raab-Traub N and Cullen BR (2006) Epstein–Barr Virus MicroRNAs Are Evolutionarily Conserved and Differentially Expressed, *PLoS Pathology*, 2, e23.
- Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, Aldler H, Rattan S, Keating M, Rai K, Rassenti L, Kipps T, Negrini M, Bullrich F and Croce CM (2002) Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia, *Proc Natl Acad Sci U S A*, 99, 15524–9.
- Calin GA, Liu CG, Sevignani C, Ferracin M, Felli N, Dumitru CD,Shimizu M, Cimmino A, Zupo S, Dono M, Dell'Aquila M, Alder H, Rassenti L, Kipps TJ, Bullrich F, Negrini M and Croce CM (2004) MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias, *Proc Natl Acad Sci U S A*, 101, 11755–60.

- Calin GA, Sevignani C, Dan Dumitru C, Hyslop T, Noch E, Yendamuri S, Shimizu M, Rattan S, Bullrich F, Negrini M and Croce CM (2004) Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers, *Proc. Natl. Acad. Sci. U. S. A,* 101, 2999–3004.
- Calin GA and Croce CM (2006) MicroRNA-Cancer Connection: The Beginning of a New Tale, *Cancer Res*, 66, 7390-7394.
- Carlson MW, Iyer VR and Marcotte EM (2007) Quantitative gene expression assessment identifies appropriate cell line models for individual cervical cancer pathways, *BMC Genomic*, 8, 117.
- Cavalli, LR and BC Liang (1998) Mutagenesis, tumorigenicity, and apoptosis: are the mitochondria involved? *Mutation Research*, 398, 19-26.
- Carrington JC and Ambros V (2003) Role of microRNAs in plant and animal development. *Science*, 301(5631), 336-8.
- Chen X (2003) A MicroRNA as a translational repressor of APETALA2 in arabidopsis flower development, *Science*.
- Chen Y, Cairns R, Papandreou I, Koong A and Denko NC (2009) Oxygen consumption can regulate the growth of tumors, a new perspective on the warburg effect, *PLoS One*, 4, e7033.
- Cho CSW (2007) OncomiRs: the discovery and progress of microRNAs in cancers, *Molecular Cancer*, 6, 60.
- Ciafre SA, Galardi S, Mangiola A, Ferracin M, Liu CG, Sabatino G, Negrini M, Maira G, Croce CM and Farace MG (2005) Extensive modulation of a set of

microRNAs in primary glioblastoma, *Biochem Biophys Res Commun*, 334, 1351–8.

- Cui X and Churchill GA (2003) Statistical tests for differential expression in cDNA microarray experiments, *Genome Biology*, 4, 210.
- Dahiya N, Sherman-Baust CA, Wang TL, Davidson B, Shih M, Zhang Y, Wood W, Becker KG and Morin PJ (2008) MicroRNA expression and identification of putative miRNA targets in ovarian cancer, *PLoS One*, 3(6), e2436.
- Dhawan P, Ahmad R, Chaturvedi R, Smith JJ, Midha R, Mittal MK, Krishnan M, Chen X, Eschrich S, Yeatman TJ, Harris RC, Washington MK, Wilson KT, Beauchamp RD and Singh AB (2011) Claudin-2 expression increases tumorigenicity of colon cancer cells: role of epidermal growth factor receptor activation, *Oncogene,* doi:10.1038/onc.2011.43.
- Edgar R, Domrachev M and Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository, *Nucl. Acids Res,* 30 (1), 207-210.
- Emery JF, Floyd SK, Alvarez J, Eshed Y, Hawker NP, Izhaki A, Baum SF and Bowman JL (2003) Radial patterning of *Arabidopsis* shoots by class III HD-ZIP and KANADI genes, *Curr. Biol,* 13, 1768–1774.
- Enright AJ, John B, Gaul U, Tuschl T, Sander C and Marks DS. (2003) MicroRNA targets in Drosophila, *Genome Biology*, 5, R1.
- Esquela-Kerscher A and Slack FJ (2006) Oncomirs — microRNAs with a role in cancer, *Nature Reviews Cancer*, 6, 259-269.

- Findley KM and Koval M (2009) Regulation and Roles for Claudin-family Tight Junction Proteins, *IUBMB Life*, 61, 431–437.
- Friedman RC, Farh KK, Burge CB, Bartel DP (2009) Most Mammalian mRNAs Are Conserved Targets of MicroRNAs, *Genome Research*, 19, 92-105.
- Goodenough DA (1999) Plugging the leaks, *Proc Natl Acad Sci USA*, 96, 319–321.
- Gordan JD, Simon MC (2007) Hypoxia-inducible factors: central regulators of the tumor phenotype, *Curr opin Genet div*, 17, 71-77.
- Greenlee RT, Hill-Harmon MB, Murray T, Thun M (2001) Cancer statistics, *CA Cancer J Clin*, 51, 15–36.
- Grimson A, Farh KK, Johnston WK, Garrett-Engele Philip, Lim LP, Bartel DP (2007) MicroRNA Targeting Specificity in Mammals: Determinants beyond seed pairing, *Molecular Cell*, 27, 91-105.
- Grün D, Wang Y-L, Langenberger D, Gunsalus KC, Rajewsky N (2005) microRNA Target Predictions across Seven Drosophila Species and Comparison to Mammalian Targets, *PLoS Comput Biol*, 1(1), e13.
- Gudi R, Melissa M.B-K, Kedishvili NY, Zhao Y and Popov KM (1995) Diversity of the Pyruvate Dehydrogenase Kinase Gene Family in Humans, *The Journal of Biological Chemistry*, 270, 28989-28994.
- Guerre-Millo M, Gervois P, Raspe E, Madsen L, Poulain P, Derudas B, Herbert JM, Winegar DA, Willson TM, Fruchart JC, Berge RK and Staels B (2000) Peroxisome proliferator-activated receptor alpha activators improve insulin sensitivity and reduce adiposity, *J. Biol. Chem*, 275, 16638-16642.

- Haeseleer DP (2005) How does clustering work, *Computational Biology,* 23, pp.1499-1501.
- Hammond, SM (2006) MicroRNA detection comes of age, *Nature Methods,* 3, 12–13.
- Harfe BD (2005) MicroRNAs in vertebrate development, *Curr Opin Genet Dev,* 15, 410–5.
- Harris AL (2002) Hypoxia—a key regulatory factor in tumour growth, Nat Rev Cancer, 2, 38–47.
- He H, Jazdzewski K, Li W Liyanarachchi S, Nagy R, Volinia S, Calin GA, Liu C-G, Franssila K, Suster S, Kloos RT, Croce CM., and Chapelle A de la (2005), The role of microRNA genes in papillary thyroid carcinoma, Proc *Natl Acad Sci USA*, 102, 19075–80.
- Heinzelmann-Schwarz VA, Gardiner-Garden M, Henshall SM., Scurry J, Scolyer RA, Davies MJ, Heinzelmann M, Kalish LH, Bali A , Kench JG,Edwards LS, Vanden Bergh PM , Hacker NF, Sutherland RL and O'Brien PM (2004) Overexpression of the Cell Adhesion Molecules DDR1 and Ovarian Cancer Claudin 3, and Ep-CAM in Metaplastic Ovarian Epithelium, *Clin Cancer Res*, 10, 4427-4436.
- Herrero J, Valencia A and Dopazo J (2000) A hierarchical unsupervised growing neural network for clustering gene expression patterns, *BMC Bioinformatics,* 17,126-136.
- Hewitt KJ, Agarwal R and Morin PJ. (2006) The claudin gene family: expression in normal and neoplastic tissues, *BMC Cancer*, 6.

- Hough CD, Sherman-Baust CA, Pizer ES, Montz FJ, Im DD, Rosenshein NB, Cho KR, Riggins GJ, Morin PJ (2000) Large-Scale Serial Analysis of Gene Expression Reveals Genes Differentially Expressed in Ovarian Cancer, *Cancer Research,* 60, 6281-6287.
- Iorio MV., Visone R, Leva GD, Donati V, Petrocca F, Casalini P, Taccioli C, Volinia S, Liu C-G, AIder H, Calin GA, Ménard S, and Croce CM (2007) MicroRNA Signatures in Human Ovarian Cancer, *Cancer Res*, 67, 869.
- Jeoung NH and Harris RA (2010) Role of Pyruvate Dehydrogenase Kinase 4 in Regulation of Blood Glucose Levels, *Korean Diabetes J*, 34(5), 274–283.
- Jiang D, Tang C, Zhang A (2004) Cluster Analysis for Gene Expression data: A Survey, I*EEE on knowledge and data engineering,* 16, 1370-1384.
- John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. (2005) Human MicroRNA targets, *PLoS Biol*, 7(3), e264.
- Kevin Chen and Nikolaus Rajewsky (2006) Natural selection on human microRNA binding sites inferred from SNP data, *Nature Genetics*, 38, 1452 – 1456.
- Kim JW, Tchernyshyov I, Semenza GL and Dang CV (2006) HIF-1-mediated expression of pyruvate dehydrogenase kinase: A metabolic switch required for cellular adaptation to hypoxia, *cell metabolism*, 3, 177–185.
- Kim VN and Nam JW (2006) Genomics of microRNA, *Trends Genet,* 22, 165–173.

- Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, Piedade I da, Gunsalus KC, Stoffel M and Rajewsky N (2005) Combinatorial microRNA target predictions, *Nature Genetics*, 37(5), 495-500.
- Krüger J, Rehmsmeier M (2006) RNAhybrid: microRNA target prediction easy, fast and flexible, Nucl. Acids. Res, 34, W451-454.
- Lai EC (2003) MicroRNAs: runts of the genome assert themselves, *Curr. Biol.* 13, R925-R936.
- Lai EC, Tomancak P, Williams and Rubin GM (2003) Computational identification of *Drosophila* microRNA genes, *Genome Biol.* 4, R42, huikm1–20.
- Lagos-Quintana M, Rauhut R, Lendeckel W and Tuschl T (2001) Identification of novel genes coding for small expressed RNAs, *Science,* 294, 853–858.
- Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T (2002) Identification of tissue-specific microRNAs from mouse, *Curr Biol,* 12, 735-9.
- Lall S, Grün D, Krek A, Chen K, Wang YL, Dewey CN, Sood P, Colombo T, Bray N, Macmenamin P, Kao HL, Gunsalus KC, Pachter L, Piano F, Rajewsky N (2006) A Genome-Wide Map of Conserved MicroRNA Targets in *C. elegans*, *Curr Biol*, 16, 460-471.
- Lau NC, Lim LP, Weinstein EG and Bartel DP (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*, *Science,* 294, 858–862.
- Lee R.C, Feinbaum R.L and Ambros V (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*, *Cell,* 75, 843–854.

- Lee R.C. and Ambros V (2001) An extensive class of small RNAs in *Caenorhabditis elegans*, *Science,* 294, 862–864.
- Lee CT, Risom T and Strauss WM (2007) Evolutionary conservation of microRNA regulatory circuits: an examination of microRNA gene complexity and conserved microRNA-target interactions through metazoan phylogeny, *DNA Cell Biol*, 26, 209-18.
- Lewis BP, Burge CB, Bartel DP (2005) Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA targets, Cell, 120, 15-20.
- Li D, Zhao Y, Liu C Chen X, Qi Y, Jiang Y, Zou C, Zhang X, Liu S, Wang X, Zhao D, Sun Q, Zeng Z, Dress A, Lin MC, Kung H-F, Rui H, Liu L-Z, Mao F, Jiang B-H and Lai L (2011) Analysis of MiR-195 and MiR-497 Expression, Regulation and Role in Breast Cancer, doi: 10.1158/1078-0432.CCR-10-1800.
- Liang P (2002) SAGE Genie: a suite with panoramic view of gene expression, *Proc Natl Acad Sci U S A*, 99(18), 11547-8.
- Liang Y, Ridzon D, Wong L and Chen C (2007) Characterization of microRNA expression profiles in normal human tissues, *BMC Genomics*, 8, 166.
- Liang Z, Wu H, Xia J Li Y, Zhang Y, Huang K, Wagar N, Yoon Y, Cho HT, Scala S and Shim H (2010) Involvement of miR-326 in chemotherapy resistance of breast cancer through modulating expression of multidrug resistance-associated protein 1, *Biochemical Pharmacology*, 79 (6), 817-824.

- Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, Rhoades MW, Burge CB and Bartel DP (1997) *C. elegans* and is regulated by the lin-4 RNA, *Cell*, 88, 637–646.
- Lim LP, Glasner ME, Yekta S, Burge CB and Bartel DP (2003) Vertebrate microRNA genes, *Science*, 299, 1540 a.
- Lin HJ, Hsieh FC, Song H, Lin J (2005) Elevated phosphorylation and activation of PDK-1/AKT pathway in human breast cancer, Br J Cancer, 93, 1372– 1381.
- Liu CG, Calin GA, Meloon B, Gamliel N, Sevignani C, Ferracin M, Dumitru CD, Shimizu M, Zupo S, Dono M, Alder H, Bullrich F, Negrini M and Croce CM (2004) An oligonucleotide microchip for genomewide miRNA profiling in human and mouse tissues, *Proc. Natl Acad. Sci. USA*, 101, 9740–9744.
- Liu H, Kohane IS (2009) Tissue and Process Specific microRNA–mRNA Co-Expression in Mammalian Development and Malignancy, *PLoS ONE*, 4(5), e5436.
- Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL , Mak RH, Ferrando AA, Downing JR, Jacks T, Horvitz HR and Golub TR (2005) MicroRNA expression profiles classify human cancers, *Nature*, 435 (7043), 834-8.
- Majer M, Popov KM, Harris RA, Bogardus C and Prochazka M (1998) Insulin downregulates pyruvate dehydrogenase kinase (PDK) mRNA: potential mechanism contributing to increased lipid oxidation in insulin-resistant subjects, *Mol. Genet. Metab*, 65, 181±186.

- Malumbres R, Sarosiek KA, Cubedo E, Ruiz JW, Jiang X, Gascoyne RD, Tibshirani R, and Lossos IS (2008) Differentiation stage–specific expression of microRNAs in B lymphocytes and diffuse large B-cell lymphomas, *Blood*, 113 (16), 3754-3764.
- Martin TA and Jiang WG (2001) Tight junctions and their role in cancer metastasis, *Histology and Histopathology*, 16(4), 1183-95.
- Maurer M, Su T, Saal LH, Koujak S, Hopkins BD, Barkley CR, Wu J, Nandula S, Dutta B, Xie Y, Chin YR, Kim DI, Ferris JS, Gruvberger-Saal SK, Laakso M, Wang X, Memeo L, Rojtman A,Matos T, Yu JS, Cordon-Cardo C, Isola J, Terry MB, Toker A, Mills GB, Zhao JJ, Murty VV, Hibshoosh H and Parsons R (2009) 3- Phosphoinositide-dependent kinase 1 potentiates upstream lesions on the phosphatidylinositol 3-kinase pathway in breast carcinoma, *Cancer Res*, 69, 6299–6306.
- Mayevsky A (2009) Mitochondrial function and energy metabolism in cancer cells: Past overview and future perspectives, *Mitochondrion*, 9 (3), 165-179.
- Megraw M, Sethupathy P, Corda B, and Hatzigeorgiou A.G. (2006) miRGen: A database for the study of animal microRNA genomic organization and function, *Nucleic Acids Research*, 35, D149-D155.
- Meltzer PS (2005) Small RNAs with big impacts, *Nature*, 435.
- Meng, F, Henson R, Lang M, Wehbe H, Maheshwari S, Mendell JT, Jiang J, Schmittgen TD and Patel T (2006) Involvement of human micro-RNA in growth and response to chemotherapy in human cholangiocarcinoma cell lines, *Gastroenterology*, 130, 2113–2129.

- Michael MZ, O'Connor SM, van Holst Pellekaan NG, Young GP and James RJ (2003) Reduced accumulation of specific microRNAs in colorectal neoplasia, *Mol Cancer Res*, 1, 882–91.
- Michaels GS, Carr DB and Askenazi M (1998) Cluster analysis and data visualization of largescale gene expression data, *Pacific Symposium on Biocomputing*, 3, 42-53.
- Morin PJ (2000) Claudin proteins in human cancer: promising new targets for diagnosis and therapy, *Cancer Res*, 65, 9603-9606.
- Murakami Y, Yasuda T, Saigo K, Urashima T, Toyoda H, Okanoue T and Shimotohno K (2006) Comprehensive analysis of microRNA expression patterns in hepatocellular carcinoma and non-tumorous tissues, *Oncogene*, 25, 2537–45.
- Nakanishi K, Ogata S, Hiroi S, Tominaga S, Aida S, Kawai T (2008) Expression of Occludin and Claudins 1, 3, 4, and 7 in Urothelial Carcinoma of the Upper Urinary Tract, *Am J Clin Pathol*, 130, 43-49.
- Oliveira SS and Morgado-Díaz JA (2006) Claudins: multifunctional players in epithelial tight junctions and their role in cancer, *Biomedical and Life Sciences*, 64, 17-28.
- Palatnik JF, Allen E, Wu X, Schommer C, Schwab R, Carrington JC and Weigel D (2003) Control of leaf morphogenesis by microRNAs, *Nature*, 20, 257–263.
- Powell DW (1981) Barrier function of epithelia, *Am J Physiol*, 241, G275–G288.

- Ramanathan A, Wang C, Schreiber SL (2005) Perturbational profiling of a cell-line model of tumorigenesis by using metabolic measurements, *Proc Natl Acad Sci USA*, 102, 5992-5997.
- Rangel LBA, Agarwal R, D'Souza T, Pizer ES, Alò PL, Lancaster WD, Gregoire L, Schwartz DR, Cho KR and Morin PJ (2003) Tight junction proteins claudin-3 and claudin-4 are frequently overexpressed in ovarian cancer but not in ovarian cystadenomas, *Clin Cancer Res*, 9, 2567–75.
- Reinhart BJ, Slack FJ, Basson M, Bettinger JC, Pasquinelli AE, Rougvie AE, Horvitz H.R. and Ruvkun G (2000) The 21 nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*, *Nature*, 403, 901–906.
- Rehmsmeier M, Steffen P, Höchsmann M, Giegerich R (2004) Fast and effective prediction of microRNA/target duplexes RNA, 10, 1507-1517.
- Reuss L (1992) Tight junction permeability to ions and water. In: Tight Junctions, edited by Cereijido M, London, CRC Press, 49–66.
- Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z,Vinsavich A, Trush V and Quackenbush J (2003) TM4: a free, open-source system for microarray data management and analysis, 34, 374-8.
- Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA and Quackenbush J (2006) TM4 microarray software suite, *Methods in Enzymology*, 411, 134-93.

- Semenza GL (2003) Targeting HIF-1 for cancer therapy, *Nat Rev Cancer,* 3, 721–732.
- Seyfried TN and Shelton LM (2010) Cancer as a Metabolic Disease, *Nutr Metab*, 7, 7.
- Seyfried TN, Mukherjee P (2005) Targeting energy metabolism in brain cancer: review and hypothesis, *Nutr Metab (Lond)*, 2, 30.
- Shin G, Kang TW, Yang S, Baek SJ, Jeong YS and Kim SY (2011) GENT: Gene Expression Database of Normal and Tumor Tissues, *Cancer Informatics*, 10,149-157.
- Shin K, Fogg VC and Margolis B (2006) Tight Junctions and Cell Polarity, *Cell and developmental biology,* 22, 207-235.
- Stine KE, On-line review: Energy metabolism and cancer.
- Sukita S, Furuse M (2000) Pores in the wall: claudins constitute tight junction strands containing aqueous pores, *J Cell Biol*, 149, 13-16.
- Swisshelm K, Macek R, Kubbies M (2005) Role of claudins in tumorigenesis, *Adv Drug Deliv Rev,* 57, 919-928.
- Takamizawa J, Konishi H, Yanagisawa K, Tomida S , Osada H, Endoh H, Harano T, Yatabe Y, Nagino M, Nimura Y, Mitsudomi T and Takahashi T (2004) Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival, *Cancer Res*, 64, 3753–6.
- Tsukita S, Furuse M and Itoh M (2001) Multifunctional strands in tight junctions. *Nature Rev Mol Cell Biol,* 2, 285–293.

- Tsukita S, Yamazaki Y, Katsuno T, Tamura A and Tsukita S (2008) Tight junction-based epithelial microenvironment and cell proliferation TJ-based epithelial microenvironment and cell proliferation, *Oncogene*, 27, 6930-6938.
- Tusher VG, Tibshirani R and Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response, *Proc Natl Acad Sci USA*, 98, 5116-5121.
- Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M, Prueitt RL, Yanaihara N, Lanza G, Scarpa A, Vecchione A, Negrini M, Harris CC, Croce CM (2006) A microRNA expression signature of human solid tumors define cancer gene targets, *Proc Natl Acad Sci U S A*, 103, 2257–61.
- Warburg O (1956) On the origin of cancer cells, *Science*, 123, 309-314.
- Watanabe T, Takeda A, Mise K, Okuno T, Suzuki T, Minami N and Imai H (2005) Stage-specific expression of microRNAs during *Xenopus* development. *Febslt*, 579 (2), 318-324.
- Wende AR, Huss JM, Schaeffer PJ, Giguere V, Kelly DP (2005) PGC-1alpha coactivates PDK4 gene expression via the orphan nuclear receptor ERRalpha: a mechanism for transcriptional control of muscle glucose metabolism, Mol Cell Biol, 25, 10684–10694.
- Wu P, Inskeep K, Bowker-Kinley MM, Popov KM and Harris RA (1999) Mechanism responsible for inactivation of skeletal muscle pyruvate dehydrogenase complex in starvation and diabetes, *Diabetes*, 48(8), 1593-9.

- Wightman B, Ha I and Ruvkun G (1993) Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans, Cell,* 75, 855–862.
- Xu P, Vernooy S.Y, Guo M and Hay BA (2003) The *Drosophila* MicroRNA Mir-14 suppresses cell death and is required for normal fat metabolism, *Curr. Biol,* 13, 790–795.
- Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, Stephens RM, Okamoto A, Yokota J, Tanaka T, Calin GA, Liu CG, Croce CM and Harris CC (2006) microRNA molecular profiles in lung cancer diagnosis and prognosis, *Cancer Cell*, 9, 189–98.
- Zhang B, Xiaoping Pan X, Cannon CH, Cobb GP and Anderson TA (2006) Conservation and divergence of plant microRNA genes, *The Plant Journal*, 46, 243–259.
- Zhang L, Volinia S, Bonome T, Calin GA, Greshock J, Yang N, Liu CG, Giannakakis A, Alexiou P, Hasegawa K, Johnstone CN, Megraw MS, Adams S, Lassus H, Huang J, Kaur S, Liang S,Sethupathy P, Leminen A, Simossis VA, Sandaltzopoulos R, Naomoto Y, Katsaros D, Gimotty PA, DeMichele A, Huang Q, Bützow R, Rustgi AK, Weber BL, Birrer MJ, Hatzigeorgiou AG, Croce CM, Coukos G (2008) Genomic and epigenetic alterations deregulate microRNA expression in human epithelial ovarian cancer, *PNAS*, 105 (1), 7004 – 7009.
- Zhang Y, Ma K, Sadana P, Chowdhury F, Gaillard S, Wang F, McDonnell DP, Unterman TG, Elam MB, Park EA (2006) Estrogen-related receptors

stimulate pyruvate dehydrogenase kinase isoform 4 gene expression, *J Biol Chem*, 281, 39897–39906.